
FuncTracker

Discovering Shared Code
(to aid malware forensics)

Presenter: Charles LeDoux
University of Louisiana at Lafayette

Shifting Focus of Malware Research

- New focus is on forensics tasks
 - Old question: *What?*
 - New questions: *Who? Why?*
-

Relationships: Putting it together

- Single instance → Single piece of the puzzle
 - Relationships indicate fitting of pieces
 - Key Relationship: Shared Code
-

Key Relationship: Shared Code

The screenshot shows the Symantec Connect website interface. At the top left is the Symantec logo and a 'Connect' link. A search bar contains the text 'Enter keywords to search...'. Below this is a navigation bar with 'COMMUNITY: Security', 'Blogs', and 'Security Response' (the active page). A 'Login or Register to participate' link is on the right. A welcome message reads: 'Welcome to the new look of Symantec Connect. Click here to find out what's changed.' The main content area features a blog post titled 'W32.Duqu: The Precursor to the Next Stuxnet', updated on 24 Oct 2011, with translations available in Japanese. The author is identified as 'Symantec Security Response' and a 'SYMANTEC EMPLOYEE'. The post has '+11' votes and '11 Votes' shown. Below the title are social sharing options: '+ Share' and 'Share this on Google+ as Arur'. The main text of the post begins with: 'On October 14, 2011, a research lab with strong international connections alerted us to a sample that appeared to be very similar to Stuxnet. They named the threat "Duqu" [dyū-kyū] because it creates files with the file name prefix "~DQ". The research lab provided us with samples recovered from computer systems located in Europe, as well as a detailed report with their initial findings, including analysis comparing the threat to Stuxnet, which we were able to confirm. Parts of Duqu are nearly identical to Stuxnet, but with a completely different purpose. Duqu is essentially the precursor to a future Stuxnet-like attack. The threat was written by the same authors (or those that have access to the Stuxnet source code) and appears to have been created since the last Stuxnet file was recovered. Duqu's purpose is to gather intelligence data and assets from entities, such as industrial control system manufacturers, in order to more easily conduct a future attack against another third party. The attackers are looking for information such as design documents that could help them mount a future attack on an industrial control facility.'

Stuxnet, Duqu, ... come from the same factory or factories

Stuxnet and Duqu were written on the same platform...by the same group of programmers.

... linked specific portions of code

Key Relationship: Shared Code

THE VERGE PRODUCTS ▾ REVIEWS ▾ FEATURES ▾ HUBS ▾ SHOW ▾ PODCAST ▾ ABOUT TIP US FORUMS ▾ Search articles & products 🔍

PREVIOUS STORY
◀ 'Halo 4: Forward Unto Dawn' vignette mourns its young soldiers' loss of...

NEXT STORY
▶ Amazon to Apple: the game starts now

WFRK SINDA ONLINE READ

The Wild West of hacking: Symantec hunts down the 'Elderwood Gang'

10 COMMENTS

By [Justin Rubio](#) on September 7, 2012 11:18pm ✉ Email 🐦 @itsTheLtr.co

```
PCI device listing ...
Bus No. Device No. Func No. Vendor/Device
0 0 2 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
0 0 29 8086 2702
```

Microsoft
less browser.
more Marvel.
Internet Explorer Try It Now

Industries:

- Automotive
- Defense
- Financial
- And more...

Linked attacks by **similarities in code**

Mapped out M.O.

Existing Approaches

- Clustering related malware
- Focus on *whole* binary comparison
 - Would miss single shared function
- Not Scalable
 - $O(n^2)$

FuncTracker:

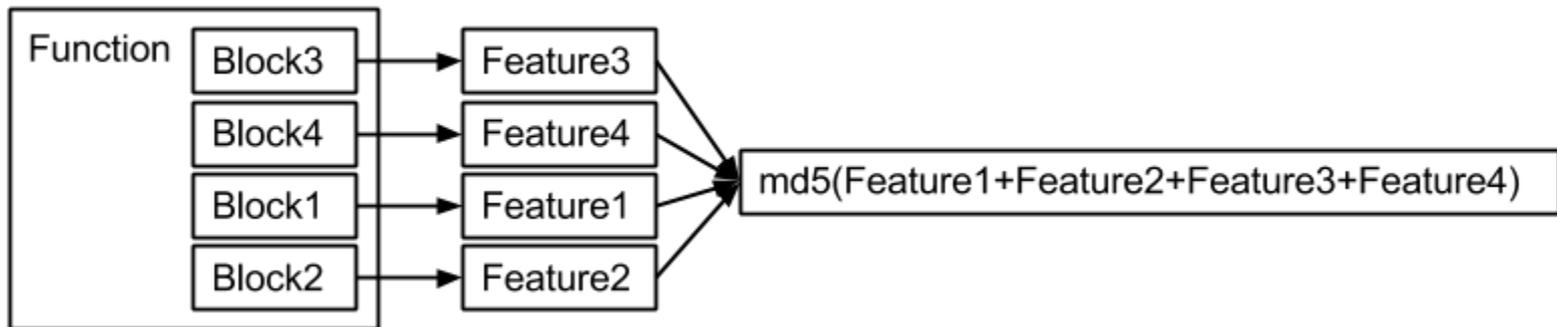
- Small, non-trivial shared code
 - Scalable
-

FuncTracker

- Granularity: Shared *Functions*
 - Whole binary comparison too coarse
 - Block level too noisy
 - Comparison: Hash Based
 - Constant time comparison
 - Syntactic and Semantic hashes
 - Exploration: Graph Based
 - Palantir intelligence platform
-

Hashes: Heart of FuncTracker

- Represent functions by set of blocks
- Represent each block by single feature
- Sort, concatenate, cryptographic hash

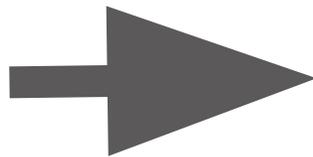


- Block features determine abstraction layer
 - BinJuice: Code, GenCode, Semantics, GenSemantics
-

Blocks: Heart of Hashes

- Code
 - Boring ol'code
 - Fragile against obfuscations
- GenCode
 - Abstract out registers and constants
 - Still fragile
 - Instruction reordering
 - Semantically equivalent substitutions

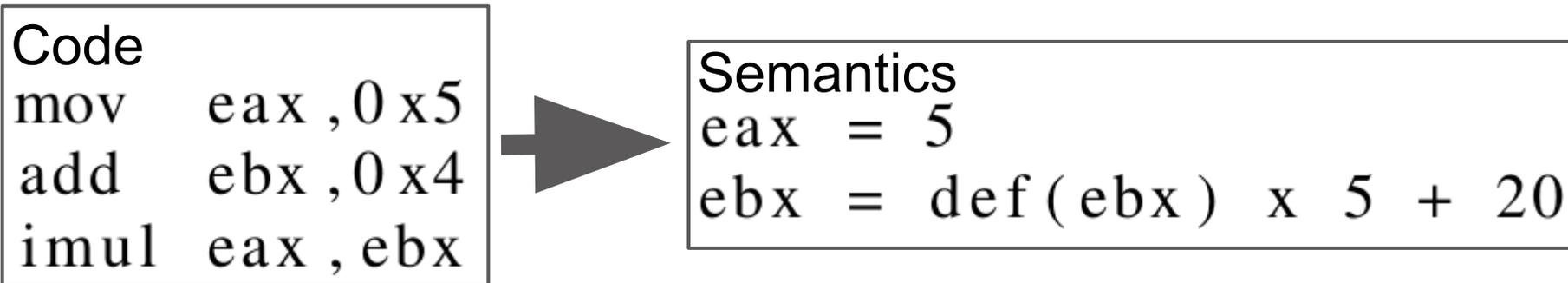
```
Code
mov    eax ,0 x5
add    ebx ,0 x4
imul   eax ,ebx
```



```
GenCode
mov    A,  N1
add    B,  N2
imul   A,  B
```

Blocks: Heart of Hashes

- Semantics
 - Effect on registers and memory
 - Symbolic interpretation
 - Algebraic simplification
 - Canonical representation



Blocks: Heart of Hashes

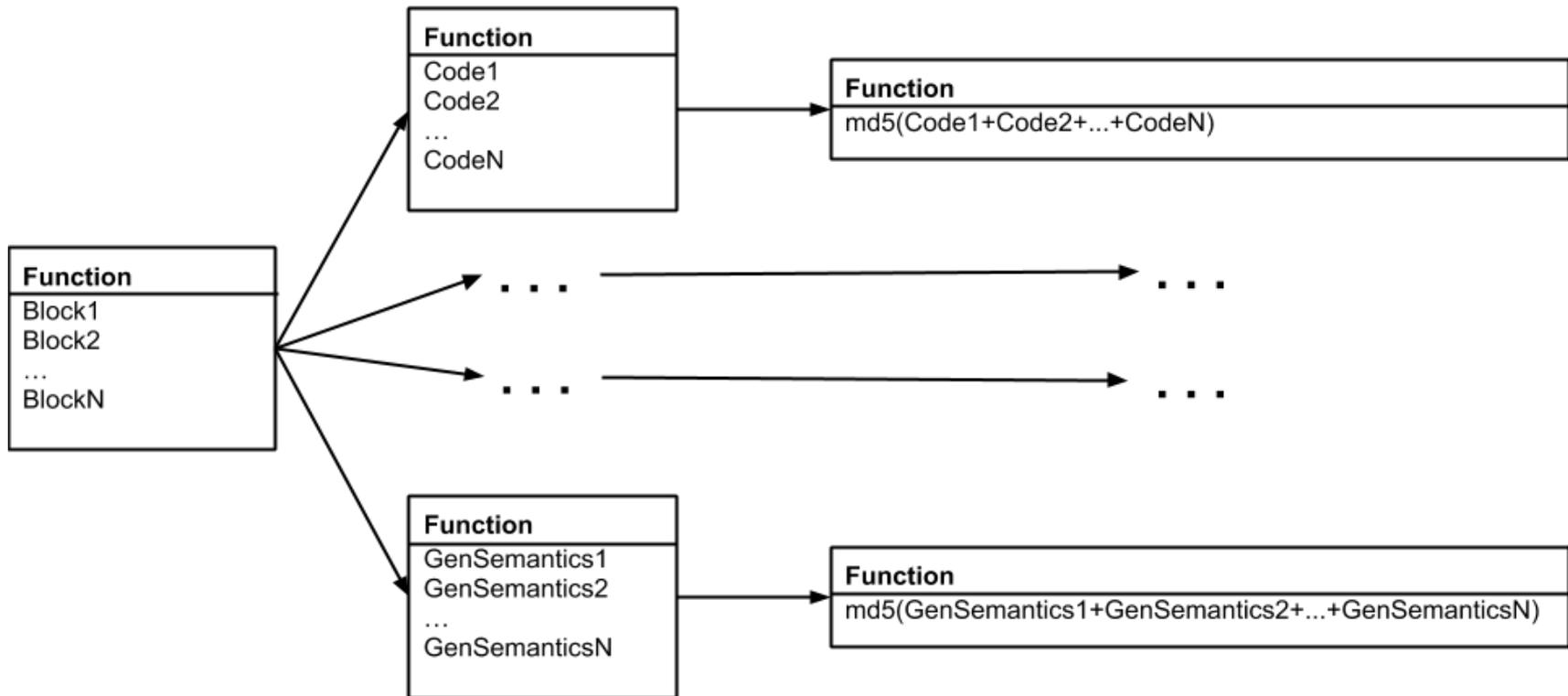
- GenSemantics
 - Analogous to GenCode

```
Semantics  
eax = 5  
ebx = def (ebx) x 5 + 20
```



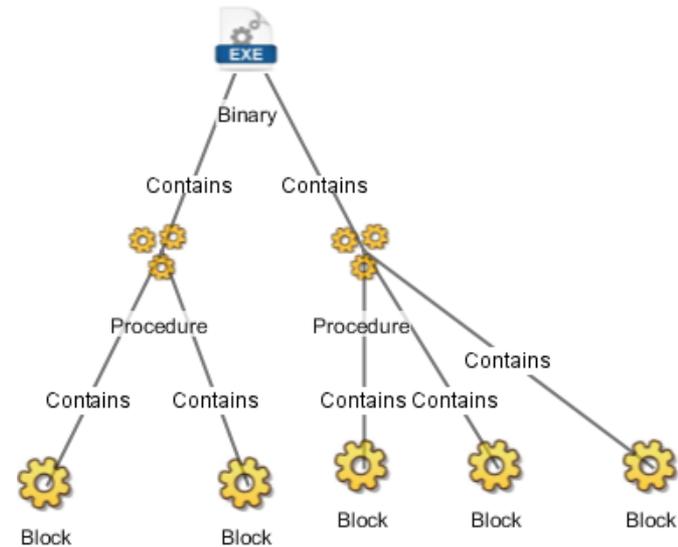
```
GenSemantics  
A = N1  
B = def (B) x N1 + N2
```

Hashes: Heart of FuncTracker



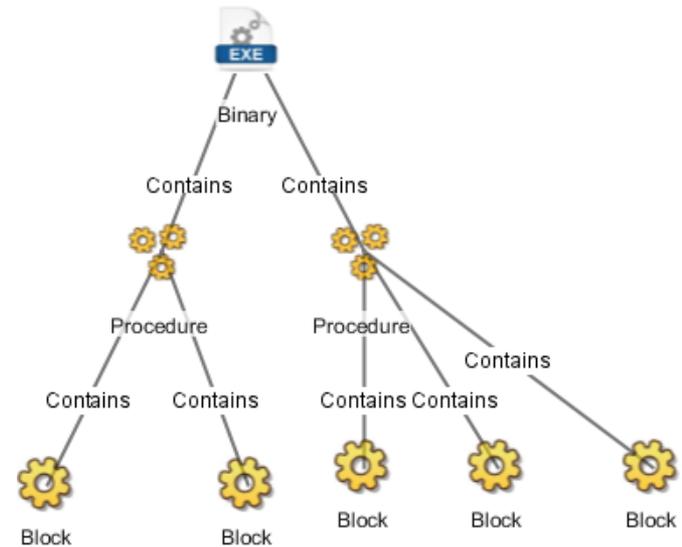
FuncTracker: Exploring Relationships

- Graph representation
- Nodes:
 - Binaries
 - Blocks
 - Functions
- Attributes:
 - Blocks: BinJuice Features
 - Functions: The different hashes
- Edges: “contains” relationship



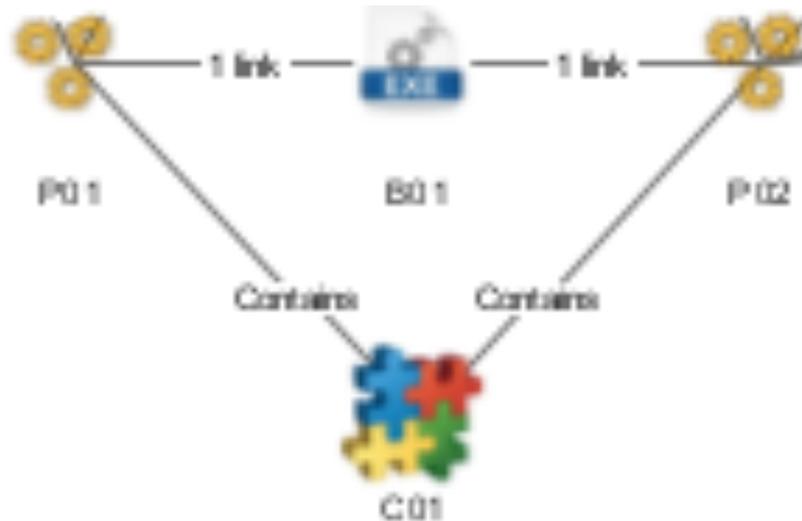
FuncTracker: Exploring Relationships

- Searches:
 - Traversal
 - Shared attribute
 - Both
- Extensible
 - Time stamp
 - Geographic location
 - Author Information
 - ...



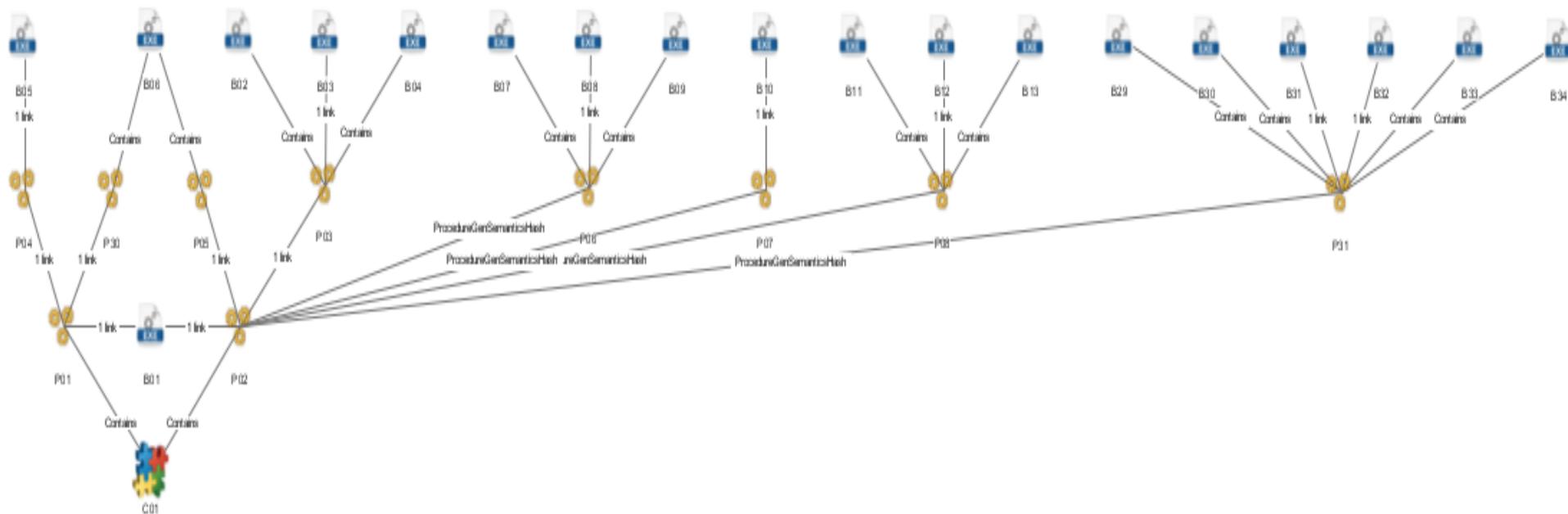
Example Use Case

- Search for shared behavior
- Start with ground truth



Example Use Case

- Search for shared behavior
- Start with ground truth
- Perform search on shared “GenSemantics”



Behavior Search Performance

	TP	FP	FN	TN
Binaries	17	1	2	90
Procedures	8	1	18	9889

What's next?

- Comprehensive evaluation
 - Extend Hashing
 - Locality Sensitive Hashing
 - Bloom Filters
-

Thank You!

Charles LeDoux
charles@charlesledoux.com
University of Louisiana at Lafayette

Arun Lakhotia
arun@louisiana.edu
University of Louisiana at Lafayette

Craig Miles
craig@craigmil.es
University of Louisiana at Lafayette

Vivek Notani
vivek200690@gmail.com
University of Louisiana at Lafayette

Avi Pfeffer
apfeffer@cra.com
Charles River Analytics
