FuncTracker

Discovering Shared Code (to aid malware forensics)

Presenter: Charles LeDoux
University of Louisiana at Lafayette

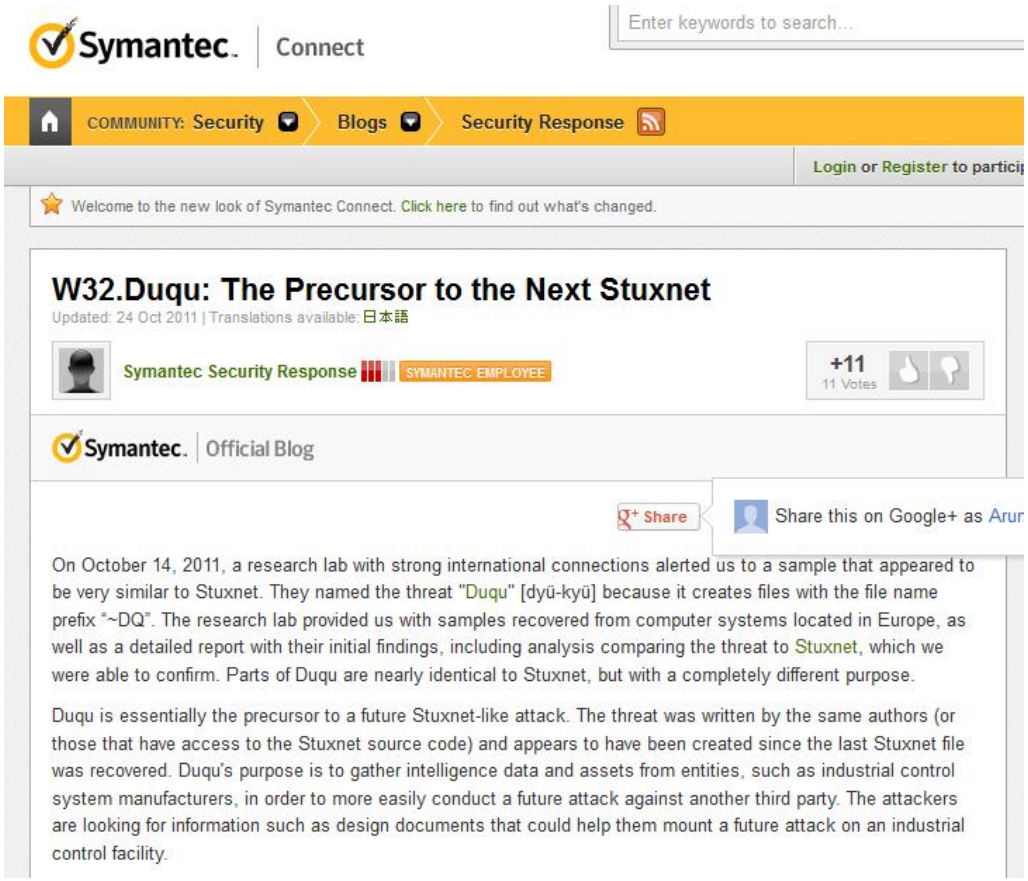
Shifting Focus of Malware Research

- 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

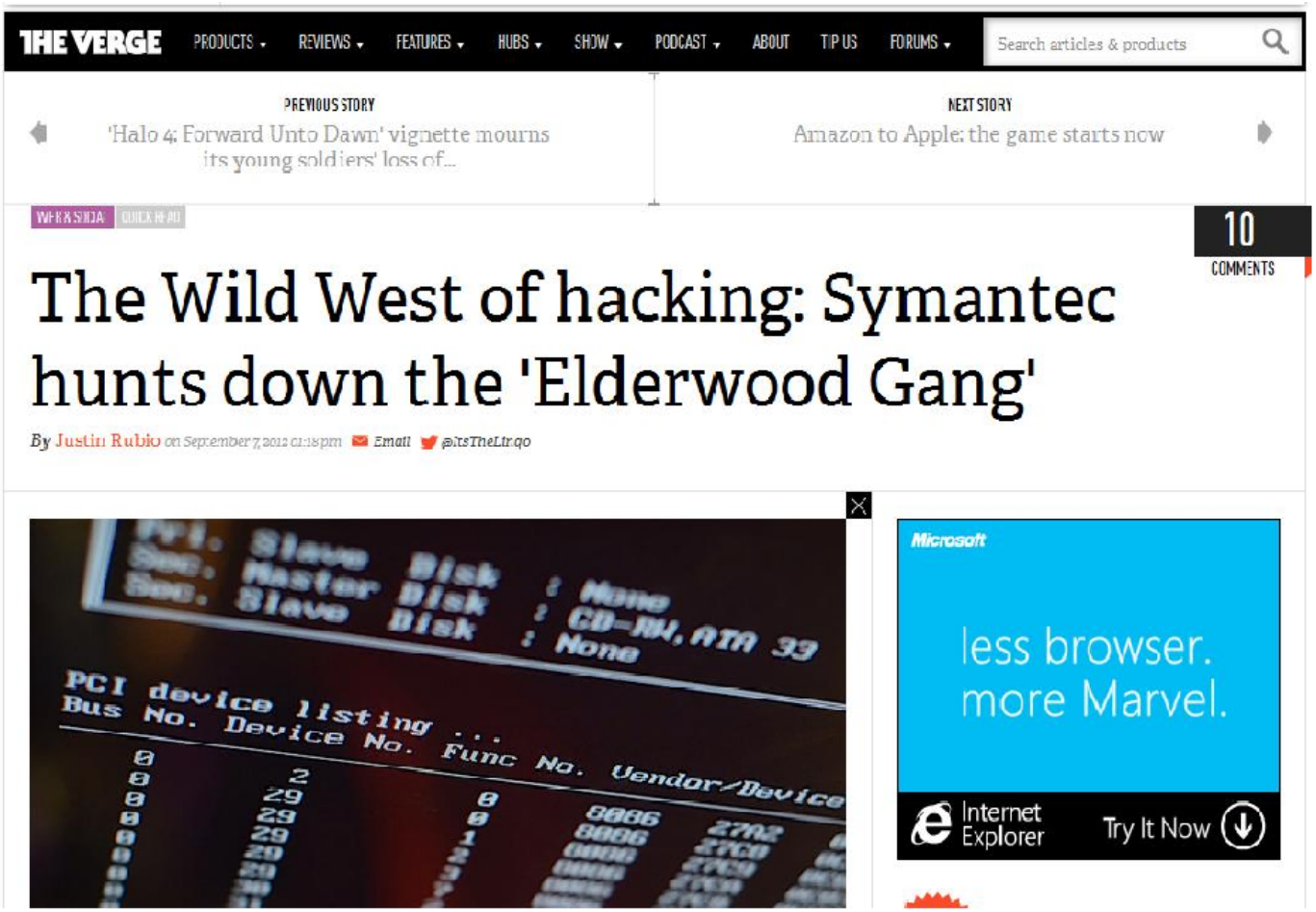


Stuxnet, Duqu, … come from the same factory or factories

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

… linked specific portions of code

Key Relationship: Shared Code



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:

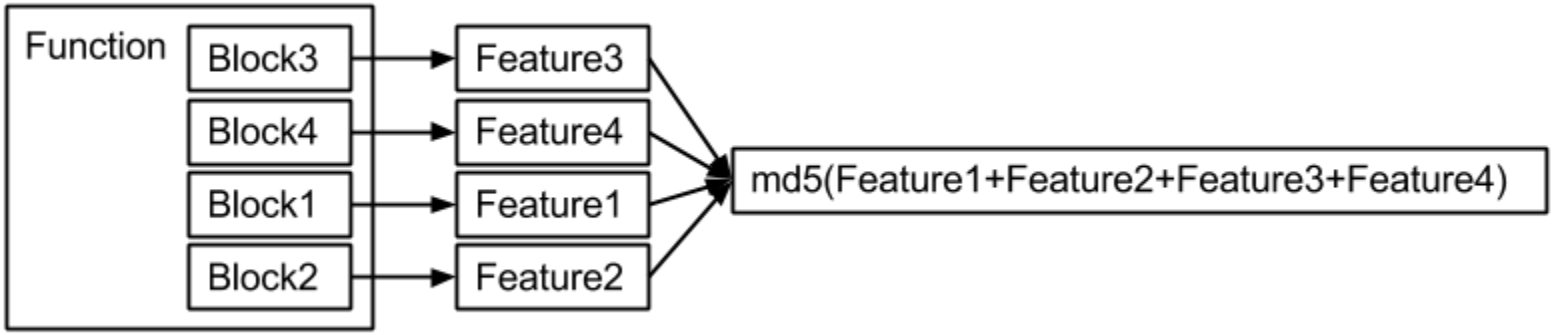
- Small, non-trivial shared code
- Scalable

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

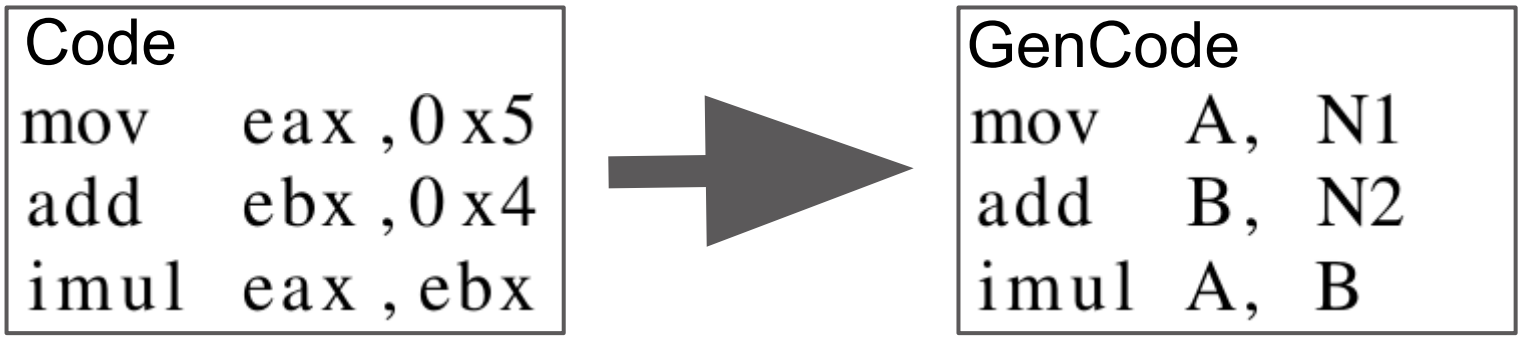
- Represent functions by set of blocks
- Represent each block by single feature
- Sort, concatenate, cryptographic hash



- 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

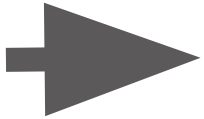


Blocks: Heart of Hashes

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

Code

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

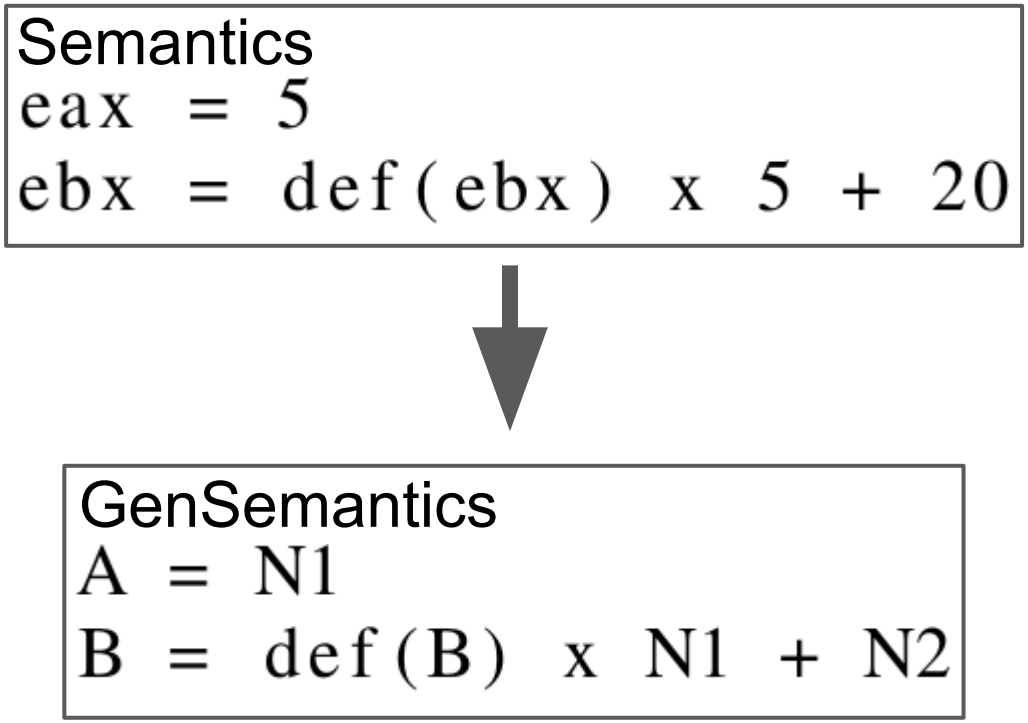


Semantics

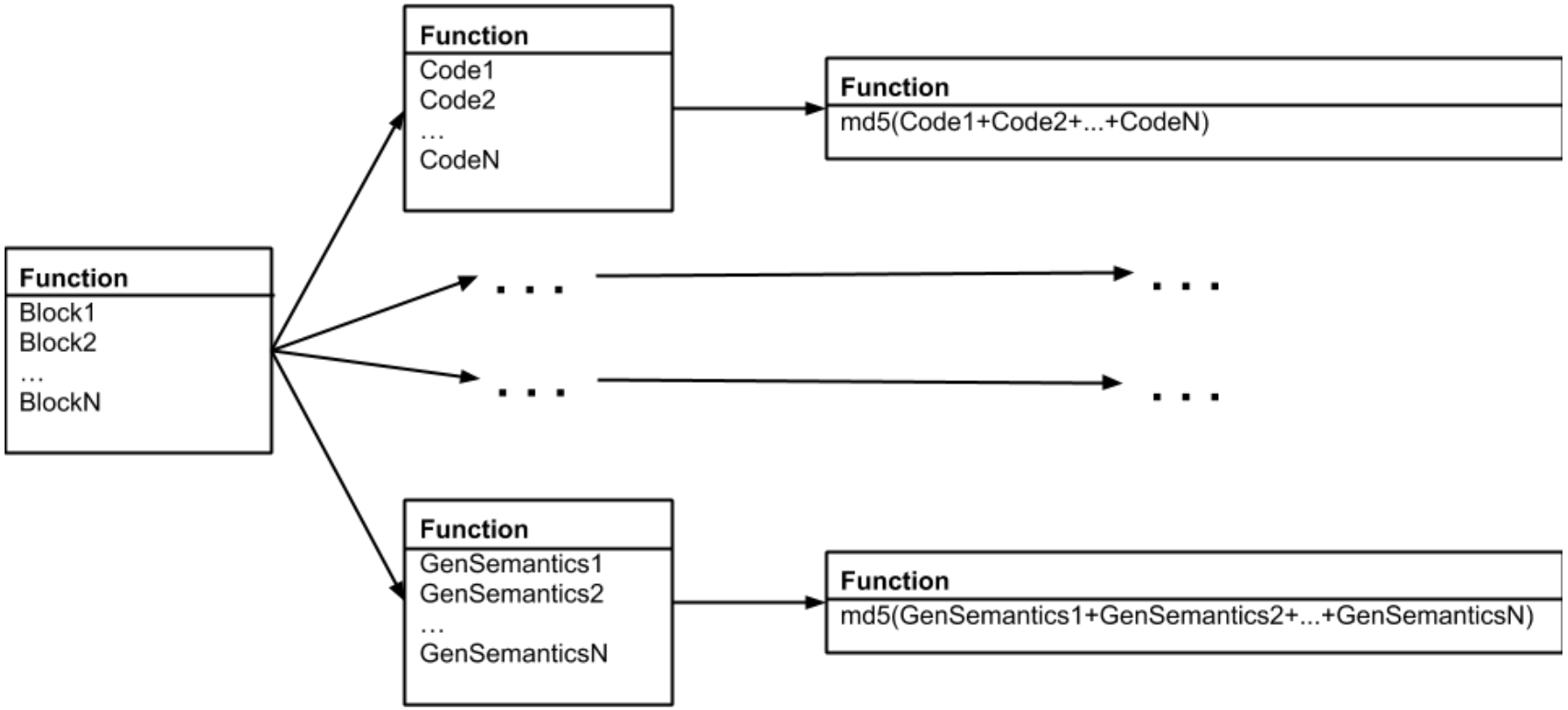
```
eax = 5
ebx = def(ebx) x 5 + 20
```

Blocks: Heart of Hashes

- GenSemantics
  - Analogous to GenCode

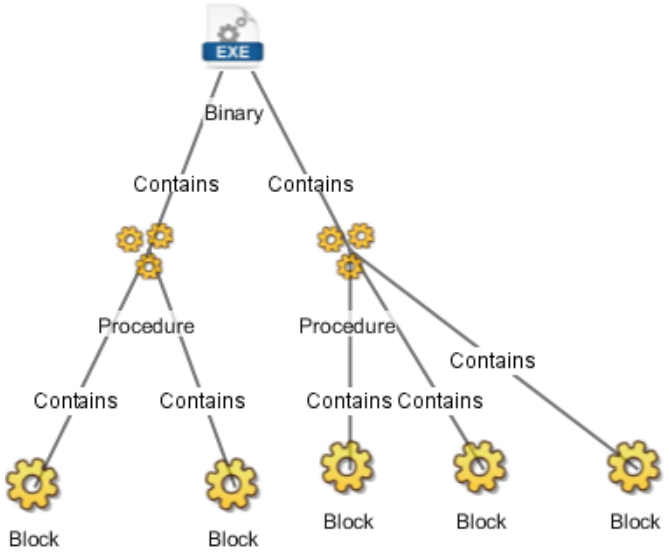


Hashes: Heart of FuncTracker



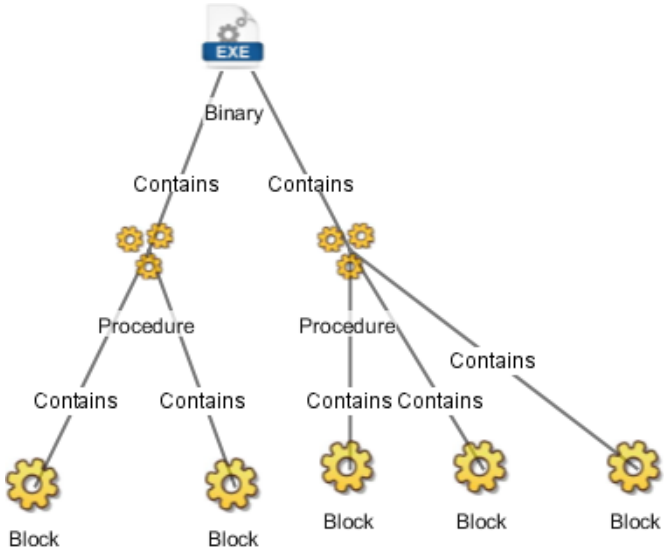
FuncTracker: Exploring Relationships

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



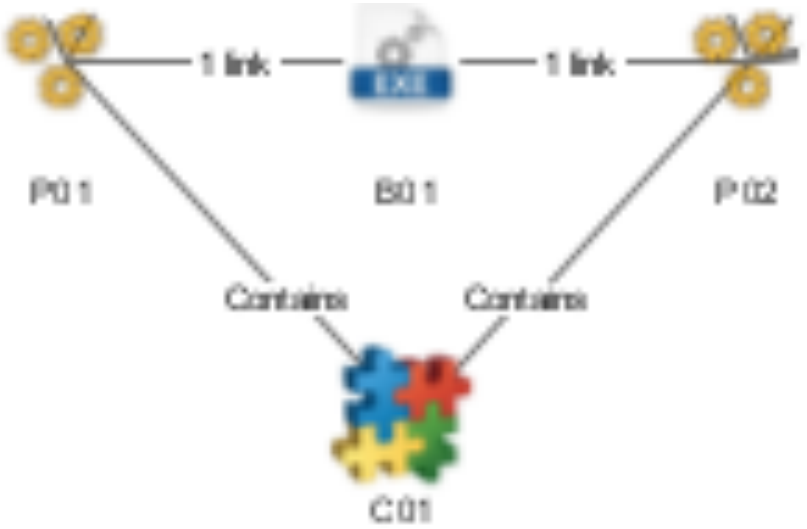
FuncTracker: Exploring Relationships

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



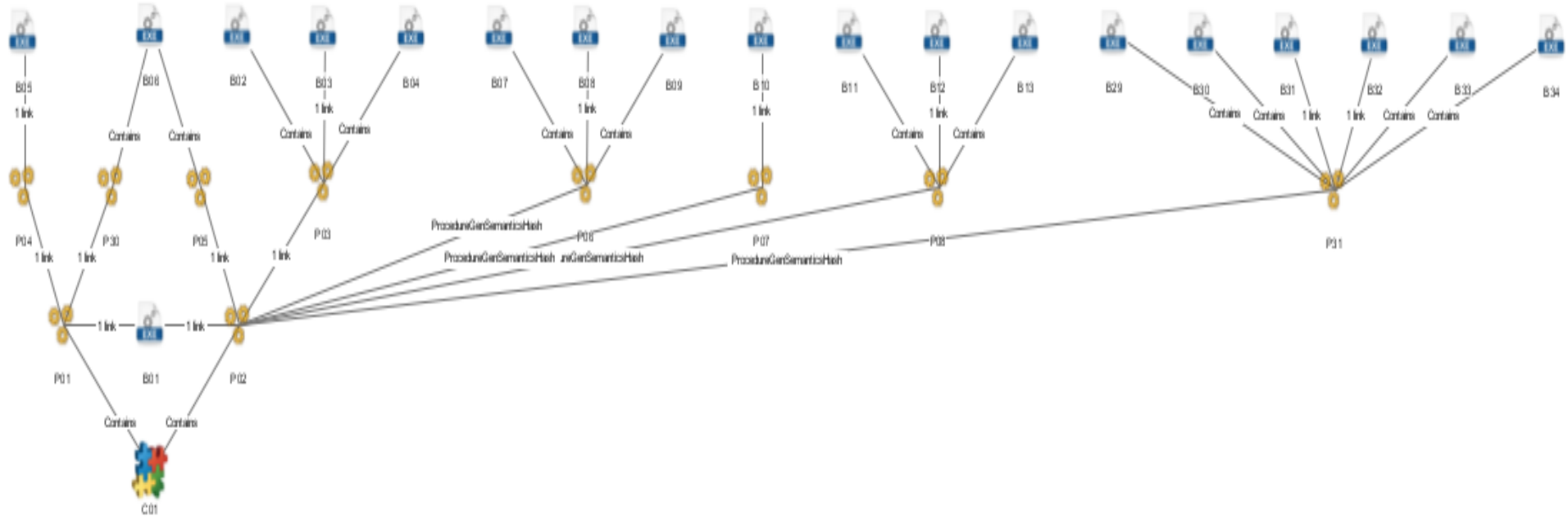
Example Use Case

- 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