Binary Diff

Atum

Text Base Binary Diff

- Key algorithm：Longest Common Subsequence(LCS) Algorithm
- Find LCS->Find Diff->Find LCS
- Pro.
  - Easy && fast
- Con.
  - Unusable on Large program

Instruction Base Binary Diff

- Disassemble->Text Diff
- Comparison of Instructions
  - Similar
  - Close
  - Negligible
  - Different

Graph Base Binary Diff

- Instruction Level && Function Level
- Make a graph
  - Vertex：Instructions, Data
  - Edge：Control Flow
- Graph Isomorphism
  - Start : Put EntryPoint && Exported Function (Function Start for function level) into queue
  - Run : Unqueue ->Instruction comparison, if True-> put next vertex into queue
  - End : Queue is empty
- Pro. Effective on finding details such as buffer size
- Con. Susceptible to compiler

Structure Base Binary Diff

1. Generate Call Graph(CG) && Control Flow Graph (CFG)
2. **Design and extract Signature from CG&CFG**
3. Get initial MAP **P** by Using **match algorithm** on CG or CFG
4. For all (a , b) ∈ **P,** Using match algorithm on **to-node sets** of a , b
5. If There is any new match add to **P**, goto 4.

Design and extract Signature from CG&CFG

- Call Graph Signature:
  - fi is a function, Sig(fi)={ai, βi, γi}
  - ai is node counts in fi (BBL counts), βi is edge counts in fi, γi is the number of functions that fi called
- Control Flow Graph Signature
  - fi is a BBL, Sig(fi)={li, Li, Si}
  - li is instruction count of fi, Li is out-degree of fi, Si is the number of functions that fi called
- Other Signature
  - Small Prime Product, Cross-references, etc.

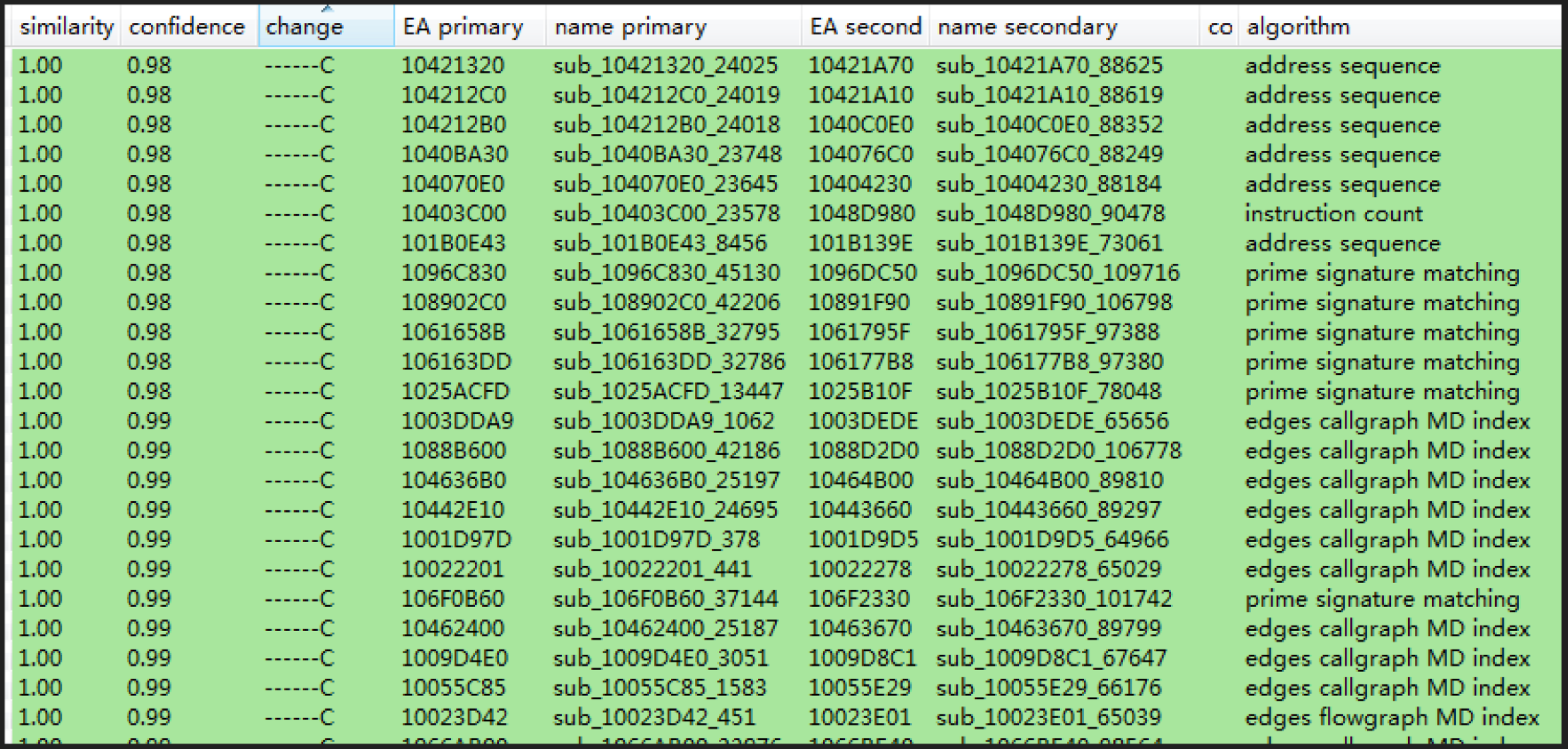
Match Algorithm

- **Let A, B is CG or CFG of two comparing binary**
- For all (a ∈ A, b ∈ B)

```
    if  sig(a)==sig(b)  &&  (∀ a’ ∈ A-a, ∀ b’∈ B-b,  sig(a’) ≠ sig(b), sig(a) ≠ sig(b’))  Then
        Add p(a)=b To P
```

Tools

- Bindiff Tool



similarity	confidence	change	EA primary	name primary	EA second	name secondary	co	algorithm
1.00	0.98	------C	10421320	sub_10421320_24025	10421A70	sub_10421A70_88625		address sequence
1.00	0.98	------C	104212C0	sub_104212C0_24019	10421A10	sub_10421A10_88619		address sequence
1.00	0.98	------C	104212B0	sub_104212B0_24018	1040C0E0	sub_1040C0E0_88352		address sequence
1.00	0.98	------C	1040BA30	sub_1040BA30_23748	104076C0	sub_104076C0_88249		address sequence
1.00	0.98	------C	104070E0	sub_104070E0_23645	10404230	sub_10404230_88184		address sequence
1.00	0.98	------C	10403C00	sub_10403C00_23578	1048D980	sub_1048D980_90478		instruction count
1.00	0.98	------C	101B0E43	sub_101B0E43_8456	101B139E	sub_101B139E_73061		address sequence
1.00	0.98	------C	1096C830	sub_1096C830_45130	1096DC50	sub_1096DC50_109716		prime signature matching
1.00	0.98	------C	108902C0	sub_108902C0_42206	10891F90	sub_10891F90_106798		prime signature matching
1.00	0.98	------C	1061658B	sub_1061658B_32795	1061795F	sub_1061795F_97388		prime signature matching
1.00	0.98	------C	106163DD	sub_106163DD_32786	106177B8	sub_106177B8_97380		prime signature matching
1.00	0.98	------C	1025ACFD	sub_1025ACFD_13447	1025B10F	sub_1025B10F_78048		prime signature matching
1.00	0.99	------C	1003DDA9	sub_1003DDA9_1062	1003DEDE	sub_1003DEDE_65656		edges callgraph MD index
1.00	0.99	------C	1088B600	sub_1088B600_42186	1088D2D0	sub_1088D2D0_106778		edges callgraph MD index
1.00	0.99	------C	104636B0	sub_104636B0_25197	10464B00	sub_10464B00_89810		edges callgraph MD index
1.00	0.99	------C	10442E10	sub_10442E10_24695	10443660	sub_10443660_89297		edges callgraph MD index
1.00	0.99	------C	1001D97D	sub_1001D97D_378	1001D9D5	sub_1001D9D5_64966		edges callgraph MD index
1.00	0.99	------C	10022201	sub_10022201_441	10022278	sub_10022278_65029		edges callgraph MD index
1.00	0.99	------C	106F0B60	sub_106F0B60_37144	106F2330	sub_106F2330_101742		prime signature matching
1.00	0.99	------C	10462400	sub_10462400_25187	10463670	sub_10463670_89799		edges callgraph MD index
1.00	0.99	------C	1009D4E0	sub_1009D4E0_3051	1009D8C1	sub_1009D8C1_67647		edges callgraph MD index
1.00	0.99	------C	10055C85	sub_10055C85_1583	10055E29	sub_10055E29_66176		edges callgraph MD index
1.00	0.99	------C	10023D42	sub_10023D42_451	10023E01	sub_10023E01_65039		edges flowgraph MD index