

# Obtaining suboperand info from Ghidra assembler v3

Goal: Get suboperand information from Ghidra’s assembler and deliver it to users via the AssemblyOperandData object

```
68 /**
69  * Construct a resolver for the given parse tree
70  *
71  * @param lang
72  * @param at the address where the instruction will start
73  * @param tree the parse tree
74  * @param context the context expected at {@code instStart}
75  * @param ctxGraph the context transition graph used to resolve purely-recursive productions
76  */
77 public AssemblyTreeResolver(SleighLanguage lang, Address at, AssemblyParseBranch tree,
78     AssemblyPatternBlock context, AssemblyContextGraph ctxGraph) {
79     this.lang = lang;
80     this.at = at;
81     this.vals.put(INST_START, at.getAddressableWordOffset());
82     this.tree = tree;
83     this.grammar = tree.getGrammar();
84     this.context = context.fillMask();
85     this.ctxGraph = ctxGraph;
86     this.operandData = AssemblyOperandData.buildAssemblyOperandDataTree(tree);
87 }
88
89 /**
90  * Resolve the tree for the given parameters
91  *
92  * @return a set of resolutions (encodings and errors)
93  */
94 public AssemblyResolutionResults resolve() {
95     AssemblyResolvedPatterns empty = AssemblyResolution.nop("Empty");
96     AssemblyConstructStateGenerator rootGen =
97         new AssemblyConstructStateGenerator(this, tree, empty);
98
99     Collection<AssemblyResolvedError> errors = new ArrayList<>();
100     Stream<AssemblyGeneratedPrototype> protStream =
101         rootGen.generate(new GeneratorContext(List.of(), 0));
102
103     if (DBG == DbgTimer.ACTIVE) {
104         try (DbgCtx dc = DBG.start("Prototypes:")) {
105             protStream = protStream.map(prot -> {
106                 DBG.println(prot);
107                 return prot;
108             }).collect(Collectors.toList()).stream();
109         }
110     }
111
112     Stream<AssemblyResolvedPatterns> patStream =
113         protStream.map(p -> p.state).distinct().flatMap(s -> {
114             // add operand data to results structure
115             empty.setOperandData(operandData);
116             return s.resolve(empty, errors);
117         });
118
119     AssemblyResolutionResults results = new AssemblyResolutionResults();
120     patStream.forEach(results::add);
121
122     results = resolveRootRecursion(results);
123     results = resolvePendingBackfills(results);
124     results = selectContext(results);
125     // TODO: Remove this? It's subsumed by filterByDisassembly, and more accurately....
126     results = filterForbidden(results);
127     results = filterByDisassembly(results);
128     results = fillMasksVals(results);
129     results.addAll(errors);
130     return results;
131 }
```

Code located at src/main/java/org/mitre/pickledcanary/assembler/sleigh/sem/AssemblyTreeResolver.java

What happens here?

What fields of AssemblyOperandData get populated here?

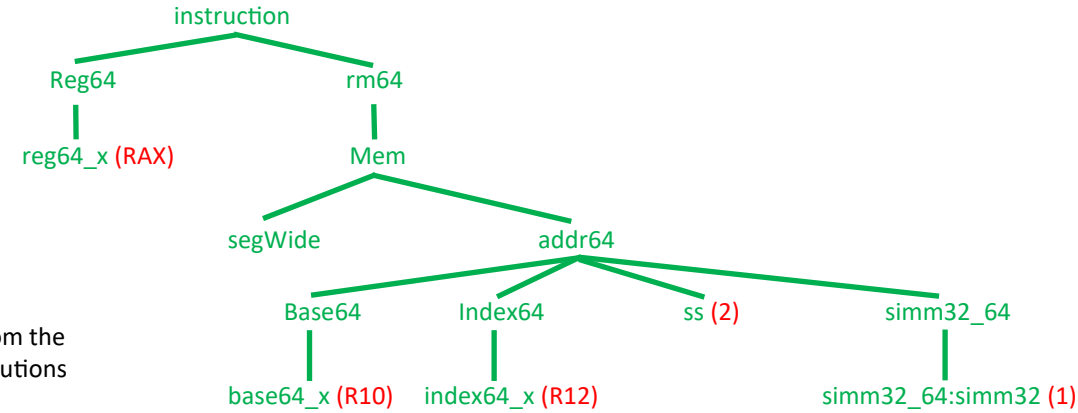
AssemblyOperandData  
buildAssemblyOperandDataTree()

5

- Build the skeleton of the operand data tree (for example, every node in the example instruction below gets an AssemblyOperandData object; basically, the structure of the tree is identical to that of the tree argument passed in)
- This AssemblyOperandData tree will eventually be returned to the user in an AssemblyResolution object

MOV RAX, qword ptr [R10 + R12 \* 2 + 1]

Brown numbers are function numbers from the Ultimate Diagram



Transfer AssemblyOperandData from the TreeResolver to the AssemblyResolutions (the results)

AssemblyOperandState  
resolve()

14

- This method is where Ghidra assembles the (sub)operands one at a time by recursively looping over the tree
- We get the AssemblyOperandData node that corresponds to the (sub)operand by checking that the code name (node names of the tree above) of the AssemblyOperandData matches that of the (sub)operand

AssemblyTreeResolver  
applyRecursionPath()

24

- Add the root shift here

AssemblyResolvedBackfill  
solve()

30

- This is where Ghidra assembles the (sub)operands for those that could not be assembled above
- We populate the ASsemblyOperandData fields for those operands here

AssemblyOperandData  
applyShifts()

32

- Add leading and trailing bytes to masks and values

AssemblyOperandData  
fillMissingMasksVals()

33

- Some (sub)operands do not have bits of their own; instead, they affect bits in the opcode
- Ghidra does not give these (sub)operands masks and values, so we give them empty byte arrays here

- i. Code name (e.g. reg64\_x)
- ii. Operand name (e.g. RAX)
- iii. Wildcard name (e.g. Q1) (if operand is wildcarded)
- iv. Wildcard index (0-based idx of wildcard in instruction (if operand is wildcarded)
- v. Operand type (register, immediate value, ...) (if operand is wildcarded) iii, iv, and v are retrieved from the assembler’s parser

- vi. Partial operand mask (leading and trailing bytes added later)
- vii. Partial operand value (leading and trailing bytes added later)
- viii. Shift (for normal operands, helps determine operand byte within the instruction)
- ix. Expression (for normal operands, helps determine operand bits within a byte and the target for jump instructions)

- x. shift (for the “instruction” code name, or root)  
\* root mask, value, and expression are not populated because opcodes cannot be wildcarded in Pickled Canary

- xi. Partial operand mask (for backfills, leading and trialing bytes added later)
- xii. Partial operand value (for backfills, leading and trailing bytes added later)
- xiii. Shift (for backfills)
- xiv. Expression (for backfills)

- xv. Operand mask (for (sub)operands that do not have any bits of their own (quasi-opcodes in previous assembler))
- xvi. Operand value (for (sub)operands that do not have any bits of their own (quasi-opcodes in previous assembler))