

```

1 #!/usr/bin/env python
2
3 import os
4 import sys
5
6 from graph_pb2 import Graph
7 from graph_pb2 import FeatureNode
8 from method_tokens import tokenize_methods_for_file, get_source_dict_graph, get_id_to_node_graph,
9 tokenize_methods_for_graph
10
11 import numpy as np
12 import string
13 import re
14
15 def filter_tokens(arr, lambda_function):
16     return [n for n in filter(lambda_function, arr)]
17
18 def condition(node, id_mapping, source_mapping, path):
19     if node == None or len(path) < 2:
20         return False
21     first = path[-2]
22     second = path[-1]
23     if ((first.contents == "VARIABLE" or first.contents == "METHOD") and second.contents == "NAME"):
24         return True
25     if (first.contents == "CLASS" and second.contents == "SIMPLE_NAME"):
26         return True
27     return False
28
29 def generate_new_name(length, b=string.ascii_uppercase):
30     d, m = divmod(length, len(b))
31     return generate_new_name(d-1, b) + b[m] if d else b[m]
32
33 def precompute_new_names(length, path):
34     with open(path, 'w') as f:
35         for i in range(length):
36             f.write(generate_new_name(i) + '\n')
37
38 def get_new_names(length, path):
39     if not os.path.isfile(path):
40         precompute_new_names(max(length, 100000), path)
41     with open(path) as myfile:
42         head = [next(myfile) for x in range(length)]
43         if len(head) < length:
44             with open(path, 'a') as f:
45                 for i in range(length - len(head) + 1):
46                     f.write(generate_new_name(len(head) + i) + '\n')
47         return [x.rstrip() for x in head]
48
49 def combine(firstId, secondId):
50     return (firstId << 16) | secondId
51
52 def get_obfuscation_names(nodeId, id_mapping, source_mapping, visited, path):
53     needs_obfuscation = set()
54     if combine(nodeId, path[-1].id) in visited:
55         return needs_obfuscation
56     node = id_mapping[nodeId]
57     if condition(node, id_mapping, source_mapping, path):
58         needs_obfuscation.add(node.contents)
59     visited.add(combine(nodeId, path[-1].id))
60     path.append(node)
61     edgeTo = source_mapping.get(nodeId)
62     if (edgeTo != None and len(edgeTo) > 0):
63         for edge in edgeTo:
64             needs_obfuscation |= get_obfuscation_names(edge.destinationId, id_mapping, source_mapping, visited, path)
65     path.pop()
66     return needs_obfuscation
67
68 def create_names_mapping(old_names_set, new_names_arr):
69     new_dict = dict()
70     index = 0
71     for old_name in old_names_set:
72         new_dict[old_name] = new_names_arr[index]
73         index += 1
74     return new_dict
75
76 def isValidSymbolMth(contents, new_names_mapping):
77     splitUp = contents.split(".")
78     length = len(splitUp)
79     last = splitUp[length - 1]
80     if (last[-2] == '(' and last[-1] == ')') and last[:-2] in new_names_mapping:
81         return True
82     for sub_path in splitUp:
83         for splt in re.split(r'\$1*', sub_path):
84             if splt in new_names_mapping:
85                 return True
86     return False

```

```

87 def middle_substitute(splitUp, new_names_mapping):
88     to_return = []
89     for sub_path in splitUp:
90         new_subpath = ''
91         index = 0
92         newSplit = re.split(r'\$1*', sub_path)
93         for splt in newSplit:
94             if splt in new_names_mapping:
95                 splt = new_names_mapping[splt]
96                 new_subpath += splt + ('$' if index != 0 else '$1')
97                 index += 1
98         to_return.append(new_subpath[:-1] if len(newSplit) > 1 else new_subpath[:-2])
99     return to_return
100
101 def substituteSymbolMth(contents, new_names_mapping):
102     splitUp = contents.split(".")
103     length = len(splitUp)
104     to_return = middle_substitute(splitUp, new_names_mapping)
105     last = to_return[length - 1]
106     if (last[-2] == '(' and last[-1] == ')') and last[:-2] in new_names_mapping:
107         to_return[length - 1] = new_names_mapping[last[:-2]] + '()'
108
109     return '.'.join(to_return)
110
111 def isValidSymbolVar(contents, new_names_mapping):
112     splitUp = contents.split(".")
113     for sub_path in splitUp:
114         for splt in re.split(r'\$1*', sub_path):
115             if splt in new_names_mapping:
116                 return True
117     return False
118
119 def substituteSymbolVar(contents, new_names_mapping):
120     splitUp = contents.split(".")
121     length = len(splitUp)
122     to_return = middle_substitute(splitUp, new_names_mapping)
123     return '.'.join(to_return)
124
125 def substitute_all(nodes, new_names_mapping):
126     for node in nodes:
127         if (node.type == FeatureNode.IDENTIFIER_TOKEN and node.contents in new_names_mapping):
128             node.contents = new_names_mapping[node.contents]
129         elif (node.type == FeatureNode.METHOD_SIGNATURE and node.contents[:-2] in new_names_mapping):
130             node.contents = new_names_mapping[node.contents[:-2]] + '()'
131         elif (node.type == FeatureNode.SYMBOL_MTH and isValidSymbolMth(node.contents, new_names_mapping)):
132             node.contents = substituteSymbolMth(node.contents, new_names_mapping)
133         elif (node.type == FeatureNode.SYMBOL_VAR and isValidSymbolVar(node.contents, new_names_mapping)):
134             node.contents = substituteSymbolVar(node.contents, new_names_mapping)
135
136
137 def obfuscate_path(path, precomputed_name_files):
138     with open(path, "rb") as f:
139         g = Graph()
140         g.ParseFromString(f.read())
141         return obfuscate_graph(g, precomputed_name_files)
142
143
144 def obfuscate_graph(g, precomputed_name_files):
145     id_mapping = get_id_to_node_graph(g)
146     source_mapping = get_source_dict_graph(g)
147     start_node = g.ast_root
148     initialPath = []
149     initialPath.append(start_node)
150     to_obfuscate = get_obfuscation_names(start_node.id, id_mapping, source_mapping, set(), initialPath)
151
152     new_names = get_new_names(len(to_obfuscate), precomputed_name_files)
153     new_names_mapping = create_names_mapping(to_obfuscate, new_names)
154     substitute_all(g.node, new_names_mapping)
155     return g
156
157 if __name__ == "__main__":
158     filePath = sys.argv[1]
159     precomputed_name_files = "precomputed_names.txt"
160     with open(filePath, "rb") as f:
161         untouched = Graph()
162         untouched.ParseFromString(f.read())
163         obfuscated_graph = obfuscate_path(filePath, precomputed_name_files)
164         before = tokenize_methods_for_graph(untouched)
165         after = tokenize_methods_for_graph(obfuscated_graph)
166
167         print("BEFORE:")
168         print(before)
169
170         print("AFTER:")
171         print(after)

```