面向软件安全的污点数据检测系统

源代码文件构成和说明

一、文件构成及每个文件的功能

a) ASTReader.cpp

系统的入口函数，扫描目标文件夹获取所有的.ast 文件，获取所有的语法树信息，并据此初始化 Callgraph、ClassTmap 和 astUnit，然后代用 checkCFG()函数来进行 block 和 stmt 层面的分析。

b) AST.cpp、AST.h

提供对 clang 生成的语法树不同类型的节点提供遍历和分析函数，通过分析不同的节点来获取所需的不同语法树信息。

（节点类型包括 FunctionDecl、CXXRecordDecl、CallExpr、VarDecl、CXXMethodDecl、FieldDecl）

c) callgraph.cpp、callgraph.h

callgraph 类数据结构，保存一个函数的各类信息，包括该函数定义结点指针、它所调用的函数和调用了他的函数指针、参数、内部变量、此函数 cfg、函数污染表、返回值得污染信息，函数内部的敏感语句的集合，函数内 block 的 IN 和 OUT 污染表信息。提供了这些信息的相关的功能操作。

d) tamp.cpp、tmap.h

CTmap 类（污染表）和 Tainted_Attr 类（污染信息）数据结构的定义。

CTmap 类保存了一系列 Vardecl*、Tainted_Attr*对（map 模板），一次保存每个变量的污染情况。Tainted_Attr 包含污染的类型（无污染 UNTAINTED，被污染 TAINTED，与某个参数相关 RELATED）和变量的类型（TYPE_VARIABLE 基类型变量，TYPE_CLASS 类类型变量，TYPE_CLASSPOINTER 类类型指针，TYPE_POINTER 指针，TYPE_UNKNOWN 未知类型）。该文件提供了污染表的合并等操作，提供了污染传播的逻辑操作（合并，覆盖等）

e) CFGtattr.cpp、CFGtattr.h

CFGInOut 类的定义，该数据结构包含 IN 和 OUT 两个污染表，分别为同一个 block 的前后的污染信息集合。该类提供了污染表相关的设置函数。

该文件还提供了对 block 信息的打印函数、IN 和 OUT 表的打印函数。还有 checkCFG 函数，该函数根据传入的 callgraph 和对应的 cfg 信息来获取 entry block，然后对该函数进行分析，生成该函数每一个 block 的 IN 和 OUT 表。

f) TCI.h

TCI（TaintedChangeInfo）指的是函数内部每一个变量的污染状态发生改变的敏感信息（包括变量 VarDecl，语句 Stmt，和改变后的污染状态信息 Tainted_Attr），该文件提供了这样的结构类型的定义。

g) TaintedStmtAnalysis.cpp、TaintedStmtAnalysis.h

提供了对 block 内部 Stmt 层面和 Expr 层面的代码分析，根据 Stmt 的类型的不同，对其造成的污染信息的传播在每个 block 的 IN 和 OUT 中修改类得以展现。所需关注的 Stmt 和 Expr 类型有复制操作、函数调用等。

h) Tout.cpp、Tout.h、tinystr.cpp、tinystr.h、tinyxml.cpp、tinyxml.h、tinyxmlerror.cpp、tinyxmlparser.cpp
功能为：系统分析过程中得到的变量的污染信息输出到结果的.xml 文件中。以上文件提供了 xml 相关和系统污染信息输出相关的功能。

二、系统的运行流程

a) 遍历目标文件夹下的所有文件，获取所有的.ast 文件，保存其路径
b) 扫描每个.ast 文件 ，遍历树上的节点，获取所需要的语法树信息
c) 用获得的语法树信息初始化 Callgraph 和 ClassTamp 的部分内容，初步获得函数调用图和每个函数的 cfg 信息。对函数递归调用报警。
d) 按调用顺序分析每个函数
   i. 如果该函数还没有被分析过，那么其 map 和 TCI_list、TCI_list_call 信息还不完整，那么分析其 cfg 内部的每一个 block，分析每个 block 内部的 Stmt 和 Expr，修改每个 block 的 IN 和 OUT 表信息，分析得到的敏感信息存入相应 callgraph 的 TCI_list 中。然后根据调用时的参数的污染情况，由其 TCI_list 和 TCI_list_call 直接确定要输出的污染的信息。
   ii. 该函数已经被分析过，那么直接使用其 map 和 TCI_list、TCI_list_call，根据参数的污染情况，输出相应的污染信息。

e) 将分析得到的变量的污染输出到 xml 文件中。