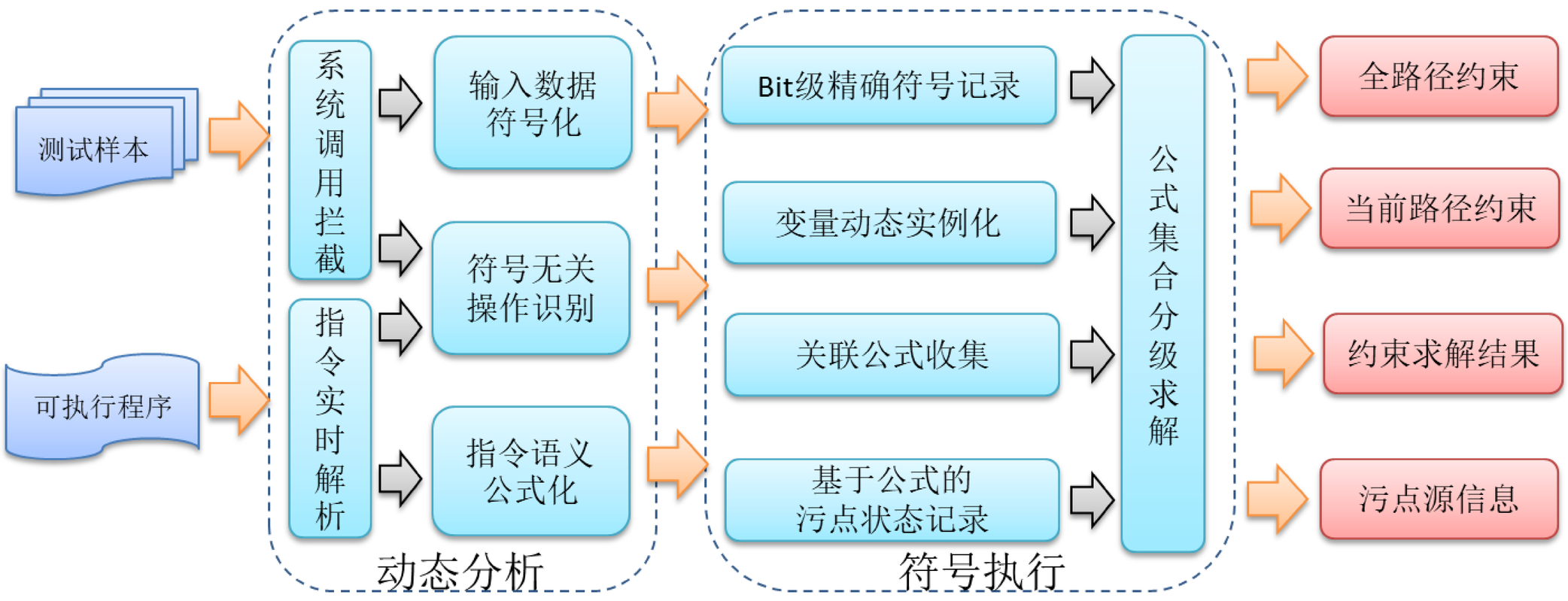
AOTA-Sym 比特级实时符号执行系统

杨轶 苏璞睿
{yangyi,purui}@iscas.ac.cn

AOTA-Sym （Application Oriented Taint Analysis-Symbolic）系统针对现有符号执行系统在二进制软件漏洞挖掘与分析工作中面临的问题研发。主要包括：1）符号化表达与推导过程精度不足，误差较大；2）基于中间语言实现符号化解析，冗余度高、分析效率低； 3）基于离线实现，空间复杂度高，且部分关键实时数据难以获取； 4）求解过程依赖于全路径约束，易被非必经节点产生的约束干扰。

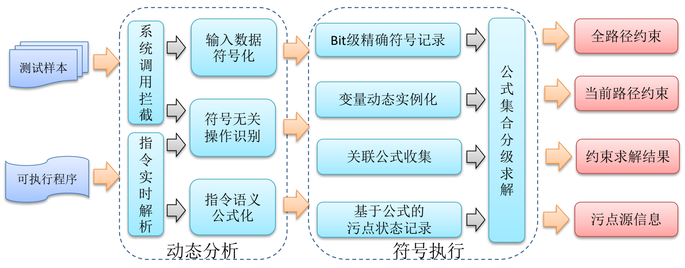
我们提出了基于偏移和位长的数据表示方法，实现了bit级精确计算与求解；提出二进制指令语义直接解析、符号无关指令识别方法，动态忽略了80%以上符号无关指令，有效提高系统效率；提出了基于公式的污点状态记录方法，在不增加复杂度的条件下，同时实现符号执行与污点传播，提高了系统分析能力；提出了基于全路径约束、当前路径约束的分级求解方法，消除了非必经节点约束干扰。



系统基于Pin插桩实现，主要包括指令动态提取、输入数据符号化、关联逻辑公式收集、公式集合分级求解、符号数据源回溯等功能，具备针对Linux、Windows等操作系统上的大型应用软件的动态逆向分析能力。系统具有如下的技术特性：

- 基于bit级符号化表示与求解，分析精度高；
- 支持命令行参数、文件、网络输入变量符号化，数据获取能力强；
- 在线符号执行，分析效率高，不受系统随机性影响；
- 支持Intel x86/x64指令集，支持Acrobat PDF、OpenSSH等大型可执行程序分析；
- 全约束求解与简化约束求解分级实现，不受前序路径中非必要约束干扰。

该系统目前已经应用于课题组研发的漏洞挖掘系统中，通过动态求解路径条件，提高程序崩溃触发能力。相关系统在3天内发现了objdump、libjpeg、libpng等广泛使用的应用和程序库中的96个未知错误。


测试样本
可执行程序
系统调用拦截
指令实时解析
输入数据符号化
符号无关操作识别
指令语义公式化
动态分析
Bit级精确符号记录
变量动态实例化
关联公式收集
基于公式的污点状态记录
公式集合分级求解
符号执行
全路径约束
当前路径约束
约束求解结果
污点源信息