软件安全入门

第六课 模糊测试

目录

CONTENTS

01 模糊测试基本原理与方法

主要内容包括模糊测试概述、基本原理与组成。

模糊测试的思想是构造所有可能的输入，并将输入传递给被测目标程序，然后监控目标程序在接收输入后是否出现异常情况，以此来发现软件中存在的缺陷和故障。

模糊测试不同于源代码审查。首先，模糊测试不需要源代码的支持；其次模糊测试会使用大量畸形数据对软件进行测试；最后，模糊测试过程通过实际执行的方式分析软件行为，能够有效地避免非直接跳转、代码混淆等因素的干扰。

模糊测试也面临着数据样本空间大、等价测试用例多、漏洞判定困难、测试对象运行环境差异显著等难题。

模糊测试系统划分为三个模块：

- ✓ **数据生成模块**：按照一定的策略生成构造测试用例，这些测试用例可能部分符合规范的输入，部分违反规范输入格式。
- ✓ **环境监控模块**：主要负责将数据生成模块生成的数据传递给测试对象并控制测试对象的运行。
- ✓ **状态监控模块**：负责监控测试对象的执行状态，得到的运行状态可以作为反馈指导数据生成模块的测试用例生成过程。

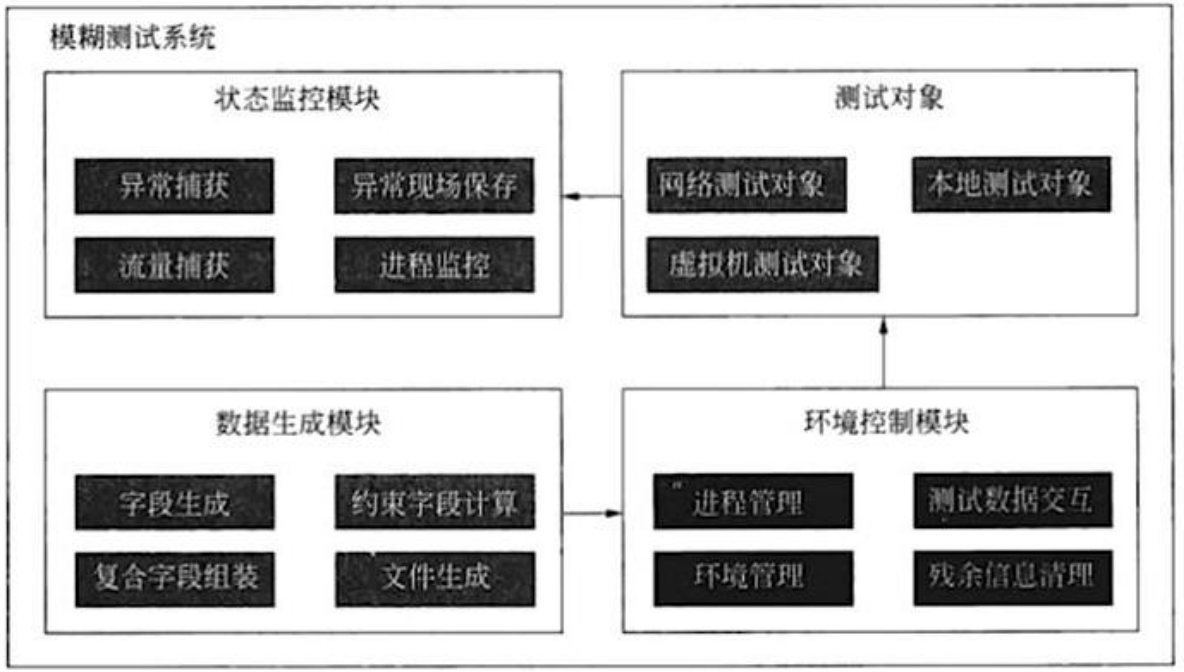

图 6-1　模糊测试系统架构

针对不同的测试对象和环境，模糊测试主要有3种工作模式：

网络模式：是指模糊测试系统仅通过网络通信的方式与测试对象进行交互。适用于测试对象是网络应用程序或主要通过网络的方式提供服务的情况。

本地模式：是指模糊测试系统在测试对象运行的操作系统上运行环境控制模块进程或代理进程。涉及到进程管理、文件管理、流量采集和软件调试等关键技术。

虚拟机模式：是指模糊测试系统主要通过虚拟机的控制接口对测试对象进行控制。在此模式下，模糊测试系统可以实现本地模式下的所有功能。

02

模糊测试基础方法

主要内容包括数据生产方法、环境控制技术、状态监控技术等。

测试数据生成的方法包含基本类型数据生成和复合类型数据生成，基本数据类型包含整型、长整型、无符号整型、字、双字与字符串等。复合数据类型是将基本数据类型进行组合。

1.基本类型数据生成方法

① **预定义序列**：包含大量用户预先定义的值，这些取值来源于用户经验以及各种收集的顺序。

② **随机序列**：在没有经验只是的前提下可采用随机序列来覆盖基础类型数据的状态空间，适用于测试数据取值为平均分布的情况

③ **幂增长**：一般来说，程序的输入、输出遍历在较小值的分布概率要远大于较大值的分布，使用幂增长的方式可以快速覆盖小数值和大数值。

2.复合类型数据生成方法

可以通过基础类型数据的顺序排列、乱序、选择、相关等方式构成复合类型数据。

2.2 环境控制技术

环境控制模块的功能是控制测试对象的运行，将生成的测试用例强制输入给测试对象，维护测试对象运行所需的环境，其中用到的技术按照功能划分为三类：

运行环境维护技术

环境控制模块对测试对象运行的环境进行控制与维护。例如测试过程对环境产生了影响，运行环境维护可采用快照备份、注册表恢复、文件恢复等技术支撑运行环境的恢复维护。

程序运行控制技术

测试程序运行在本地系统下对程序的启动、暂停、调试、修改、终止等控制。例如环境控制模块可以通过Windows系统内置的CreateProcess系列的API系统调用创建测试对象进程等。

数据强制输入技术

测试数据输入的方式根据测试对象及测试环境的不同分为网络、文件、用户操作等多种方式。数据强制输入技术因此也分为网络数据输入技术、文件数据输入技术和用户操作输入技术以及内存数据修改技术。

2.3 状态监控技术

模糊测试系统在执行模糊测试时需要对测试对象的生命周期、执行状态、异常状态与输入输出进行监控，主要目的是判断测试对象是否出现异常，指导环境控制模块对测试环境与测试对象进行更准确的控制，指导数据生成模块生成更有效率的测试数据。根据监控类型的不同可分为：

生命周期监控技术

判断软件是否处于正常的运行状态下以及判断程序是否出现异常。例如通过获得测试对象的进程句柄来判断进程的运行状态。

输入输出监控技术

程序的输入输出主要通过异步过程与系统提供的I/O设备进行交互，对程序的输入输出进行监控简单易行，方法通用性较好，因此得到了广泛应用。

执行状态监控技术

当模糊测试系统发现了测试对象发生崩溃时，测试状态的执行状态信息将有助于对程序崩溃进行初步筛选。如程序崩溃时栈的数据结构、寄存器状态、异常链等信息。

03

模糊测试优化方法

主要内容灰盒模糊测试、混合符号执行、基于反馈的模糊测试。

灰盒模糊测试方法是通过逆向分析程序二进制代码和输入数据的标准格式，生成有针对性的违背数据格式规范的测试数据，从而提高模糊测试的效率。其主要功能原理图如下：

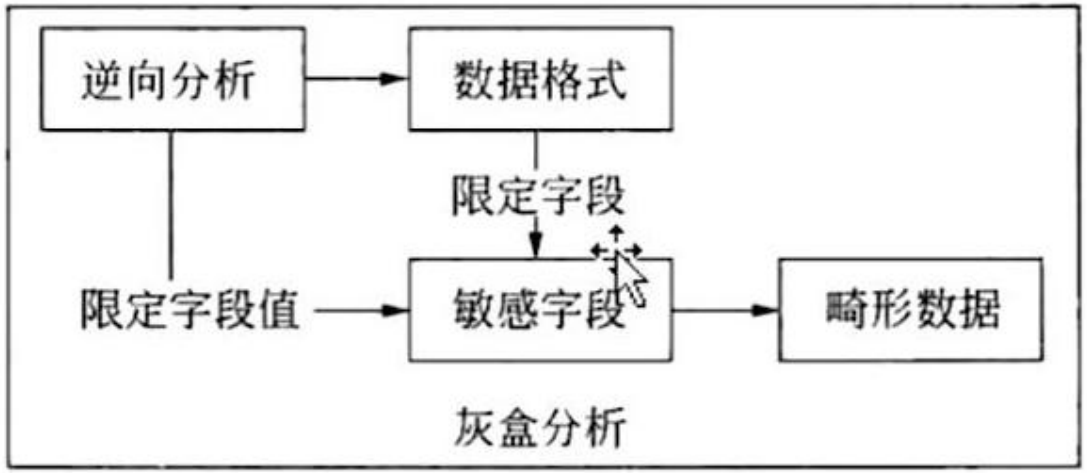

图 6-11　灰盒模糊测试功能原理

白盒模糊测试是一种结合了模糊测试和符号执行的方法。白盒模糊测试的过程如下：

通过静态分析能够获得输入中敏感字段的位置，而字段的取值除了遍历和随机取值外，还可以通过符号执行的方法进行优化。通过分析程序指令，将程序中内存与寄存器的值表示为输入变量的表达式，然后联立每个分支语句所代表的约束表达式，再用符号执行技术求出程序执行各路径分支的一个满足条件的测试用例。通过这些测试用例，模糊测试系统可以更多地覆盖程序地各个代码分支，在减少测试用例地同时提高代码覆盖率。

模糊测试过程是一个循环的反复测试过程，通过统计分析前序测试用例特征与测试结果特征，可以指导后续测试数据的生成，这称为基于反馈的模糊测试方法。

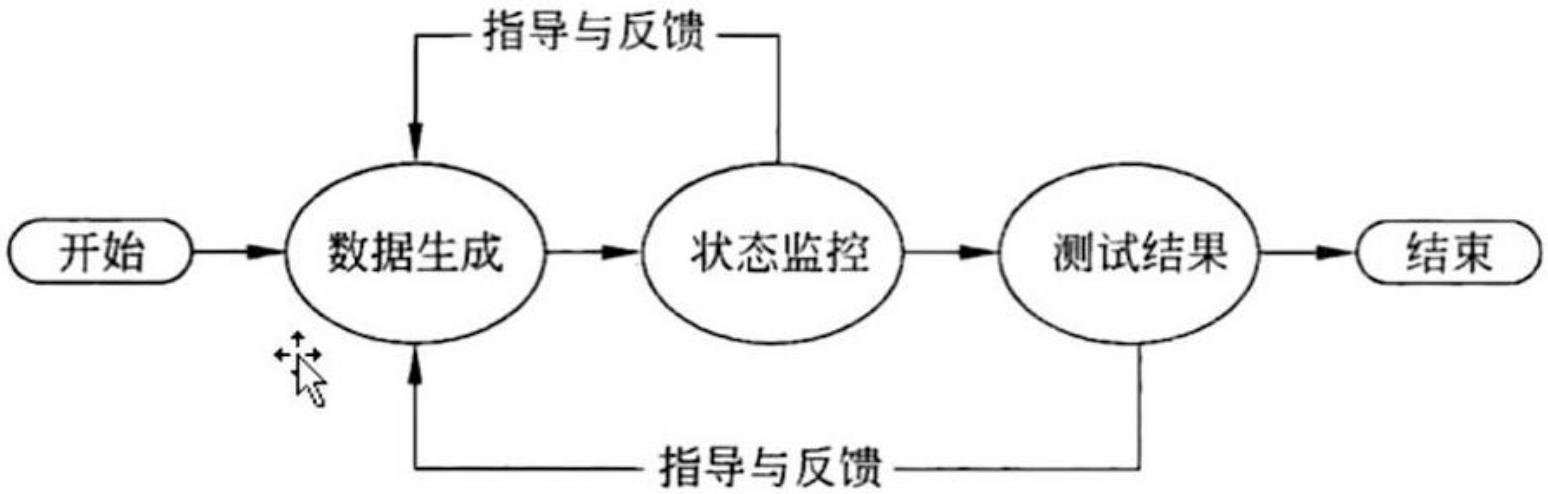

图 6-13　基于反馈的模糊测试流程

04 分布式模糊测试

主要内容包括分布式控制结构、分布式模糊测试策略、动态适应机制等。

分布式控制的基本功能是保证分布式节点不会大量重复的工作，在此基础上根据分布式节点处理能力的区别分配工作，并根据节点动态的变化进行自适应的调整。

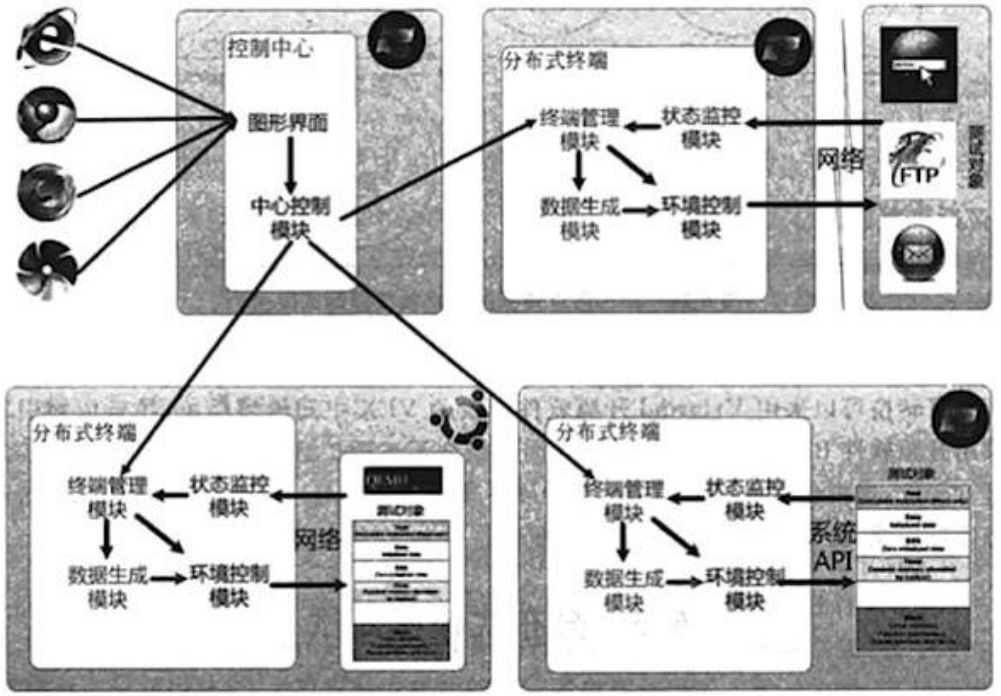

图 6-15　分布式模糊测试架构

分布式模糊测试要求各分布式终端执行不同的测试用例，但是应该为各分布式终端制定统一的策略文件，主要原因是如果每个分布式终端采用不同的测试策略，当分布式终端的数量由于故障、硬件资源变化等原因发生变化，就要为每个分布式终端生成新的策略文件。

因此，一种可行的分布式模糊测试策略是采用统一描述，然后各分布式终端根据自己的序号生成与其他分布式终端不同的测试用例。一个简单的方法是，每个分布式终端生成完全一样的测试用例序列，然后选择序号模终端数量的值与节点序号相等的测试用例输入给测试对象。这种方法简单易行，效率也可接受。

分布式模糊测试策略提出了按执行速度的比例划分测试用例的方法，在此基础上对其进行改进，将整个模糊测试过程划分为许多阶段，每完成一个阶段的测试，控制中心会计算根据各分布式终端的执行速度计算其占有测试用例的权值，并在此时加入新的计算资源或剔除已失效的计算资源。

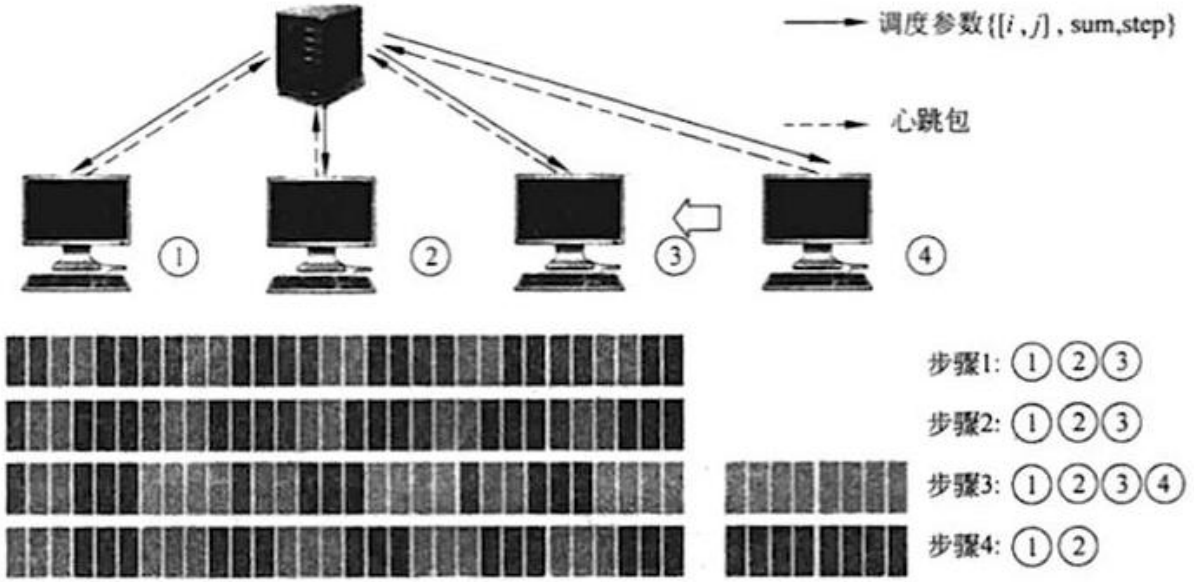

图 6-16　动态适应的分布式调度机制

05 典型工具与案例

主要内容包括几种常见的开源模糊测试工具和代表性工具Peach和Sulley的介绍。

下面列举几个常见的开源模糊测试工具：

产品名称	特点描述
ANTIPAPSER	一个用Python语言编写的API，被设计用来帮助生成随机数据
Dfuz	基本组成包括数据、函数、列表、选项、协议以及变量。模糊测试引擎对这些不同的组成部分定义的规则集进行解析以生成和传输数据。
SPIKE	能自动地更新每个字段的值，就好像实施了不同的变异操作一样。包含大约700个可诱发错误的启发式攻击列表。
Peach	是一个采用Python语言编写的跨平台的模糊测试框架，提供了一些基本的构建以创建新的模糊器，包括生成器、转换器、协议、发行器以及群组。
GPF	通过一些模式来提供相关功能，不同模式针对不同应用场景。
Sulley	一个模糊器开发和模糊测试框架，它是由多个可扩展的构件组成的，目标是不仅简化数据的表示，而且简化数据的传输以及对目标的监视。

Peach是一个功能强大的模糊测试工具，它提出了一种称为Pit的脚本文件，可以与种子文件一同使用，对文件等复杂数据有很好的描述能力。

Peach的基本功能如图所示：

- **Publisher**：是测试数据生成模块，负责根据Pit文件与种子文件创建测试用例。
- **Fuzzer引擎与代理管理模块**：是环境控制模块，负责管理测试对象及其运行环境，并与测试对象进行交互。
- **Logger**：是状态监控模块，通过Peach内置的探针提取测试状态信息。

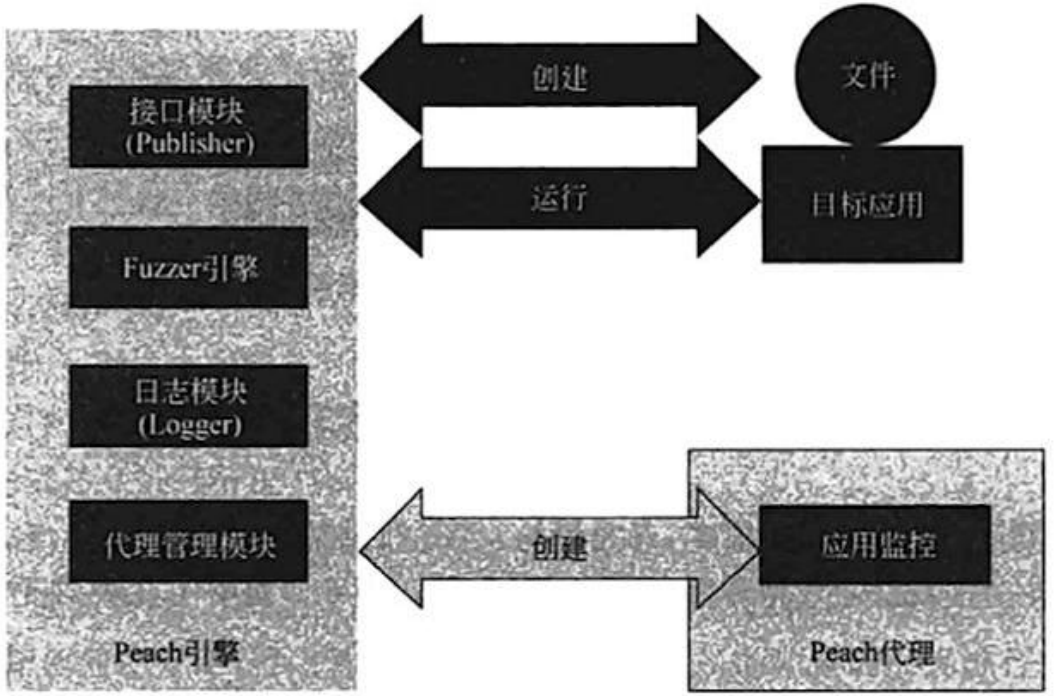

图 6-17　Peach 的基本结构

Sulley是一个分布式模糊测试框架，它通过RPC协议实现了一套分布式控制机制，同时给出了一些基础数据结构的数据生成方法，并在一定程度上支持复杂数据结构的数据生成。

Sulley的基本结构如图所示：

- **Primitives和Logos模组**：定义了各基础数据类型的数据生成方法。
- **Blocks**：定义了复杂数据类型的表示方法与数据生成方法。
- **Session管理器**：对图进行解析，并与测试对象进行数据交互。
- **代理**：提供了一些对分布式终端性能、网络流量、进程进行监控的方法。

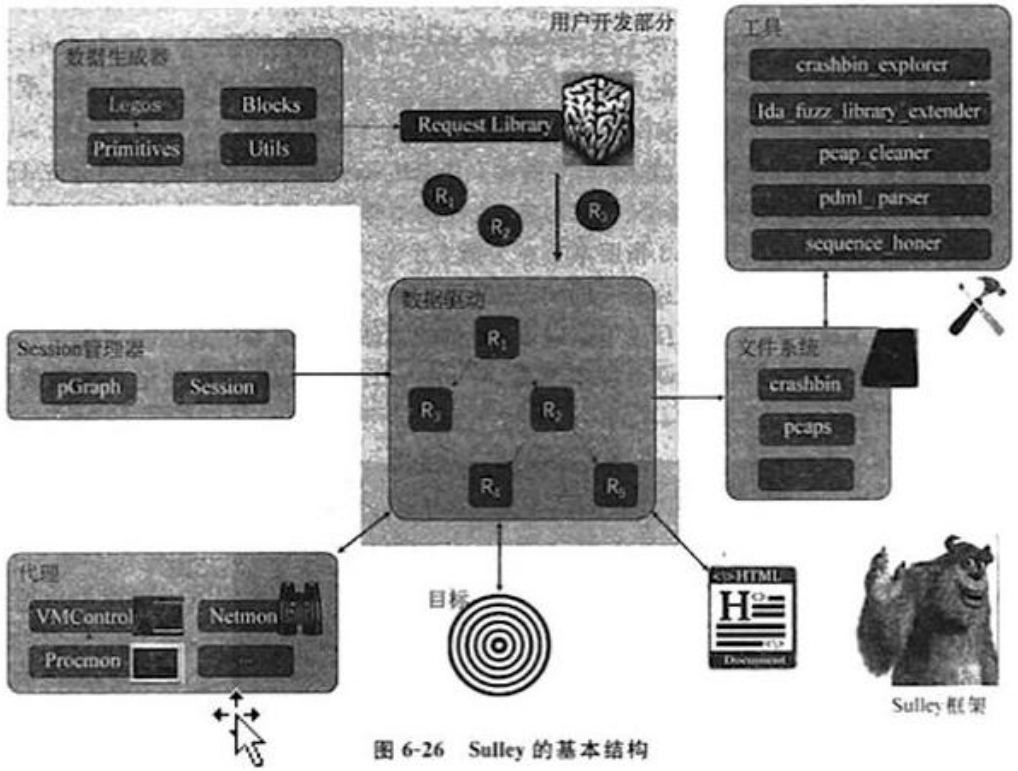

图 6-26 Sulley 的基本结构

课程小结

- **1.模糊测试基本原理**
- 主要内容包括：模糊测试概述、基本原理与组成。
- **2.模糊测试基础方法**
- 主要内容包括：数据生产方法、环境控制技术、状态监控技术等。
- **3.模糊测试优化方法**
- 主要内容包括：灰盒模糊测试、混合符号执行、基于反馈的模糊测试。
- **4.分布式模糊测试**
- 主要内容包括：分布式控制结构、分布式模糊测试策略、动态适应机制等。
- **5.典型工具与案例**
- 主要内容包括：几种常见的开源模糊测试工具和代表性工具Peach和Sulley的介绍。

谢谢观赏