

### QCON 全球软件开发大会 【北京站】2016

使用静态分析技术找到"真正"的代码质量缺陷与安全漏洞

韩葆 (Bob Han), Synopsys Software Integrity Group, 13311307163,Bao.Han@synopsys.com

**International Software Development Conference** 

# このり2016.10.20~22上海・宝华万豪酒店

### 全球软件开发大会2016

[上海站]



购票热线: 010-64738142

会务咨询: qcon@cn.infoq.com

**赞助咨询:** sponsor@cn.infoq.com

议题提交: speakers@cn.infoq.com

在线咨询(QQ): 1173834688

团・购・享・受・更・多・优・惠

**优惠(截至06月21日)** 现在报名,立省2040元/张

### HeartBleed Bug





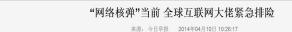
近日,安全协议OpenSSL哪出本年度最严重的安全漏洞 "心脏出血"。利用该漏洞,属客坐在自己家里电脑 射,就可以实时我取到惯多https开头网边的用户登录账号密码,包括网根、知名购物网站。电子邮件等信息, 这相当于信用卡信息被公开的程度。分分钟很关几十万的节奏呀。据360网站安全检测平台对国内120万家经过 授权的网站扫描,其中有11440个网站主机受该漏洞影响。

目前。国内大量网络包括阿黑巴巴、特н邻多个大型互联网层系被被过度微重作。已经够复了场DemSSL漏 调,而中国金融认证中心则发文表示,网根受到的影响较少。U图可以放心使用。这个漏洞为什么称为"心脏出 血"漏洞呢它的危害性有多大呢到底什么是OpenSSL呢?下面为你全面解读。









"网络核弹"当前,全球互联网大佬紧急排险

专家建议, 网站未打补丁前别急着修改密码

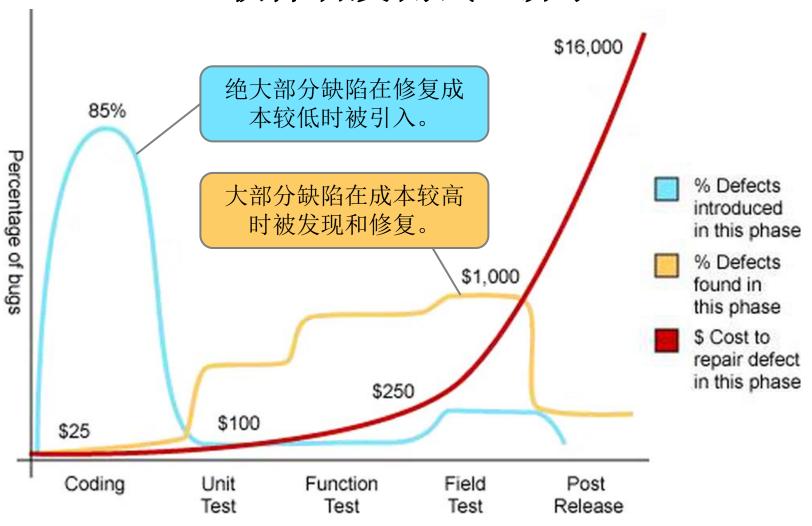
本周二,黑客和"白帽"(使用正当手段优化网站的IT人士)们经历了一个不眠之夜。

他们有的在狂欢,逐个进入戒备森严的网站,耐心地收集泄漏数据,拼凑出用户的明文 密码;有的在艰苦升级系统,统计漏洞信息,还要准备说服客户的说辞,让他们意识到问题的严 重性......

当天,从亚马逊到雅虎,多个国际大牌互联网公司都召开紧急会议,应对最新发现的互联网漏洞"心脏出血(Heart bleed)"。这是发现者们给这个号称本年度最严重安全漏洞起的形象

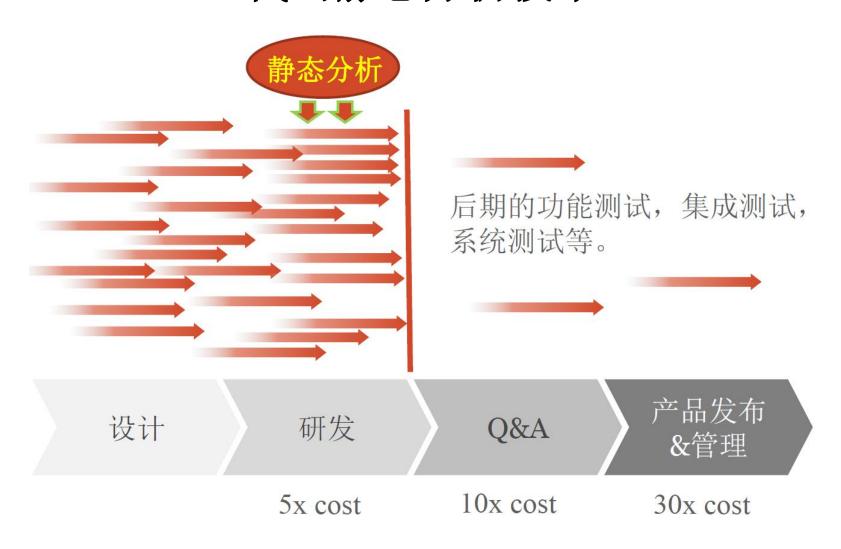


### 软件研发测试经济学



Source: Applied Software Measurement, Capers Jones, 1996

### 代码静态分析技术



### 代码静态分析

- 定义: 在不执行计算机程序的条件下,对源代码进行分析,找出代码 缺陷
- 执行方式: 一般配合静态程序分析工具进行
- 采用技术:数据流分析、机器学习、语义精简...
- 可检测类型: 死锁, 空指针, 资源泄露, 缓冲区溢出, 安全漏洞, 竞 态条件...
- 优点:
  - 能够检测所有的代码级别的可执行路径组合,快速,准确
  - 直接面向源码,分析多种问题
  - 在研发阶段开始找到并修复多种问题,节省大量时间/人力成本
- 注意:静态分析不是万能的,测试是持续的过程,非一劳永逸

### 现存问题

- 编译器警告: 保证类型安全
  - 最初级的静态分析,检测规则简单
- 中间语言分析: 检测字节码(Byte Code)的缺陷,将其重新映射到真实代码中
  - 在转换与映射过程中易造成精度丢失
- 高误报率: 目前静态分析产品的误报率普遍在30%以上。
- 缺陷种类较少,找到的问题级别不高:多数为代码规范或低级缺陷,非实际Bug
  - 如命名规范、类定义规范,最佳实践.....
- 易用性较低:基本上都是一次性的使用工具,无法与SDLC集成
  - SCM集成:如SVN,CVS,Perforce,Git
  - Bug Tracking: 如Bugzilla,Jira

### 改进型的静态分析方案

- 基于Meta Compilation的静态分析:
  - 由斯坦福大学教授Dawson Engler提出, 在深度理解代码与程序语义的基础上检测缺陷
  - 旨在查找"真正的代码缺陷"
- 实现原理:
  - 使用可扩展的metal语言定义正确性Checker
  - 将程序的源码使用状态机进行抽象描述(State Machine Abstraction)。
  - 使用xgcc系统匹配Checker与抽象状态机状态,找到问题所在的点。
- 可准确检测实际的Bug(内存和指针问题、资源泄露、缓冲区溢出,数组越界,心脏出血漏洞...)
  - 能够检测高达亿行级别的代码库,避免"状态爆炸"
  - 使用模型检验与符号执行技术, 误报率降低至15%以下
- 算法已步入实际应用
  - 面向企业的Coverity 软件
  - 面向开源代码的Coverity SCAN

### 源码分析-数据流分析

- 源码分析可以探知开发者的想法: "x=1"需要在调用 "do something"后继续执行。
- 提出警告: if循环没有包含所有语句

### 如何进行Java代码静态分析?

Java语言被编译成JVM bytecode - 在运行时被转换成本地可执行 代码的分析

### 选项一

- 分析 byte-code: 用户编译他们的软件, 然后分析编译后的可执行文件与调试信息, 分析引擎联系找到的缺陷与源代码位置
- 某些开源工具的实现原理

### 选项二:

- 获取所有的Java编译过程并执行分析
- Bytecode分析工作仍旧存在,但包含更多的内容

### 基本的工作流

- 获取所有编译过程
- 每当"javac(或其他相关API)"被调用后,编译获取系统记录所有的编译器选项,操作,源代码与调用的库文件
- 面向源代码和库文件可进行全面编译后分析
- 找到的缺陷将被展示给研发人员修复

### 如何分析缺陷?

- 过程间分析(Intra-procedural analyses)将考虑每一个合理的可执行路径
  - 快速修剪不可行路径是一件很麻烦的事情!
  - 数学方案
- 获取一系列的函数定义
  - 资源分配
  - 调用....
- 过程间分析
  - Bytecode 分析将创建函数定义

### 如何分析缺陷?

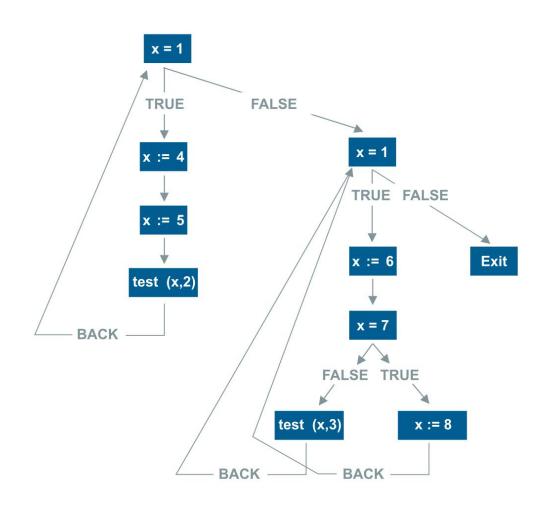
- 数据流分析将跟踪 所有应用中的不可信数据
  - "source"
  - "sink"
  - 二者之间必须进行验证
- 某些使用智能静态分析, 例如:
  - checked this return value for null 19 times out of 20
  - accessed this field under a lock 19 times out of 20
  - called base.Foo() in 19 overrides of Foo() out of 20

### 找到潜在Bug其实只是难题之一

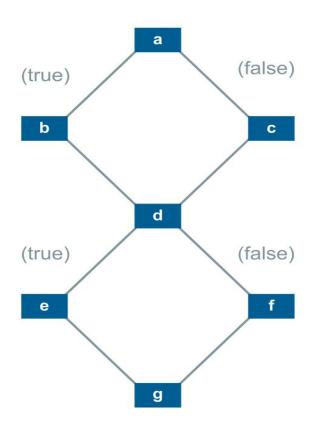
- •消除误报非常难
- 将复杂的缺陷解释出来很难
- 只找潜在的一次性缺陷是很难的

### 难题!

### **Control Flow Graph**

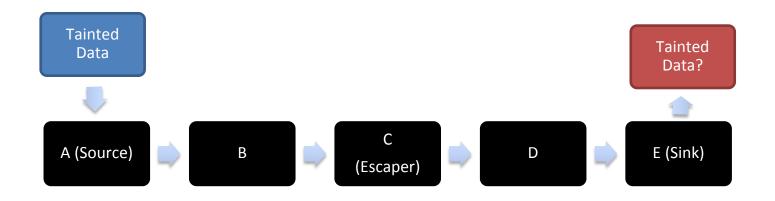


### 如何简化搜索?



### 数据流分析

### Advanced data flow



- 动态"fuzzing"
- 基于上下文分析

### Coverity OWASP top 10: JSP&ASP

	OWASP 10 - 2013	CWE映射
A1:	注入	77, 78, 88, 89, 90, 564, 917
A2:	失效认证与会话管理	259, 321, 384, 798
A3:	跨站脚本攻击 (XSS)	79, 80, 81, 82, 83, 84, 86, 87
A4:	不安全的直接对象引用	22, 23 ,36
A5:	安全配置错误	4, 7, 86, 650
A6:	敏感信息泄露	321
A7:	功能级访问控制缺失	425, 862, 863
A8:	跨站请求伪造	352
A9:	使用含有已知漏洞的组件	NA
A10	: 未验证的重定向和转发	938

### Java 缺陷

### Web 应用安全缺陷 (OWASP Top 10)

- 跨站脚本攻击
- SQL 注入
- 命令行注入
- 路径遍历...

#### 资源泄露

- 数据库连接资源泄露
- 资源泄露
- Socket & Stream 泄露

#### 并发数据访问异常

- 变量非原子更新
- 双重检查锁定
- 数据竞态条件
- Volatile非原子更新
- Servlet 属性无效锁定
- 单例模式竞态条件

#### 程序假死

- 线程死锁
- 死锁

### 空指针引用

- Null检查后引用空指针
- 直接引用返回的空指针
- Null检查前引用空指针

#### API 使用错误

- 无效迭代器使用
- 不可修改的集合错误
- 己释放资源调用

#### 性能缺陷

- 低效率方法使用
- 在循环中连接字符串
- 冗余同步

#### 逻辑错误

- 不可达代码
- 未使用变量
- 常量表达式
- 非本地资源不当使用
- 整数溢出
- 不当分号

### Java 缺陷

### 类层次结构不一致

- 调用 super.clone() 或 supler.finalize()失败
- 父函数调用丢失
- 构造函数中使用虚函数

#### 控制流缺陷

- 在Finally模块中返回
- Switch语句中break丢失

#### 错误处理缺陷

- 未验证的返回值

### 数据库操作

- 不正确的实体哈希
- Load函数返回值错误验证
- 不完全持续周期
- get()不当使用

### 代码可维护性缺陷

- 调用已过期方法
- 显式垃圾收集
- 非静态方法中设置静态变量
- 复制/粘贴错误
- 不可达代码

#### 可疑代码

- 参数次序错误
- 格式错误

### C# 缺陷 Powered by Eric Lippert

#### 资源泄露

- 数据库连接资源泄露
- 资源泄露
- Socket & Stream 泄露

#### API 使用错误

- 已释放资源调用

#### 并发数据访问异常

- 变量非原子更新
- 数据竞态条件

#### 性能缺陷

- 低效率方法使用
- 在循环中连接字符串
- 冗余同步

#### 程序假死

- 线程死锁
- 死循环

#### 可疑代码

- 复制/粘贴错误
- 参数次序错误
- 格式错误

### 类层次结构不一致

- 调用 base.close() 或 base.dispose()失败
- 父函数调用丢失

#### 控制流缺陷

- 可疑的额外分号
- 不一致比较
- 不兼容的类型比较

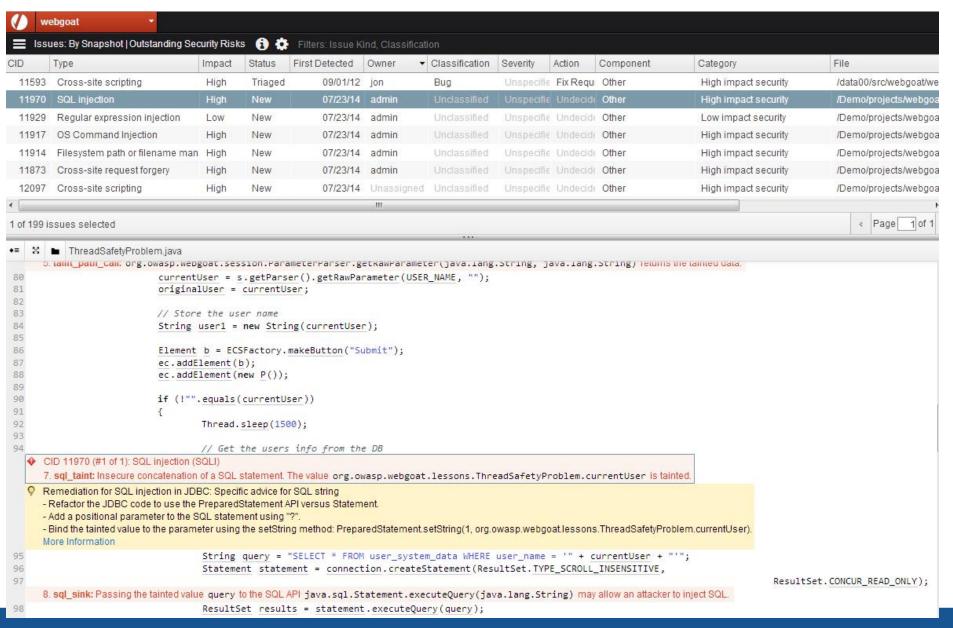
### 空指针引用

- Null检查后引用空指针
- 直接引用返回的空指针
- Null检查前引用空指针

### 算术错误

- 错误移位操作
- 不正确的表达式
- 表达式计算过程中溢出

### 检测实例-SQL Injection



### Copy-paste error in real-world code

```
if (returns!=null) {
  r.retvals = ScopeParser.parseTypedArgList(returns,
returns.getText(), g.tool.errMgr);
  r.retvals.type = AttributeDict.DictType.RET;
 r.retvals.ast = returns;
if (locals!=null) {
  r.locals = ScopeParser.parseTypedArgList(locals,
locals.getText(), g.tool.errMgr);
 r.locals.type = AttributeDict.DictTYpe.LOCAL;
 r.locals.ast = returns;
```

### C语言静态分析

Checker描述(metal 语言) 检测代码:

```
#include "extend-lang.hpp"
 START_EXTEND_CHECKER( basic_leak, int store );
 ANALYZE TREE ()
={
     // Do not unmangle
     // We want the actual C functions; with unmangling we could get
     // e.g. Foo::malloc
     CallSite malloc( "mallog", /*unmangle*/false );
     CallSite free ( "free", /*unmangle*/false );
     LocalVar lv;
     AnySubpart lv sp(lv);
     if( MATCH( lv sp = malloc( ) ) ) {
         SET STATE ( lv sp, 1 );
         ADD EVENT( lv_sp, "alloc", "Allocation assigned to " << lv_sp ); }
     else if ( MATCH ( free ( lv sp ) ) ) {
         CLEAR STATE ( lv sp ); }
 ANALYZE END OF PATH ()
      const ASTNode *t;
     int v;
     FOREACH IN STORE (t, v) {
          COMMIT ERROR( t, "leaked", t << " was not freed before the end of path" );}
 END EXTEND CHECKER();
 MAKE MAIN ( basic leak )
```

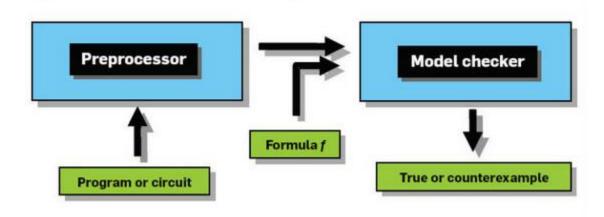
7

}

```
#include<stdlib.h>
1
       test()
(1) Event alloc: Allocation assigned to "p"
Also see events: [leaked]
       void *p = malloc(10);
      /* ... */
       //free(p);
                 "p" was not freed before the
(2) Event
leaked:
                 end of path
Also see events: [alloc]
```

### XGCC系统

- 符号执行
  - 不执行程序,用符号值表示程序变量的值,模拟程序执行
  - 可以分析代码的所有/部分语义信息
  - 避免状态爆炸
- 模型检验



### C/C++ 缺陷-Part 1

#### 资源泄露

- 内存泄露
- Microsoft COM 内存泄露
- Object资源泄露
- 不当delete

### 未初始化变量

- 返回语句丢失
- 未初始化的指针/标量/数组 读写
- 类或结构体中未初始化的数据成员

### 并发缺陷

- 死锁
- 竟态条件(Race conditions)
- 阻塞调用误用

### 算术错误

- 负变量不当使用
- 异常符号扩展
- 整数溢出
- 除零异常

### 内存崩溃

- 内存访问溢出
- 字符串长度计算错误
- 缓冲区溢出
- 写指针溢出
- 负数组索引写入
- 内存错误分配
- 错误的内存释放

### 非法内存访问

- 不正确的delete操作
- 溢出指针读取
- 越界读取
- 返回指针至本地变量
- 负数组索引读取
- 已释放指针读/写
- 不兼容的指针转换

### 控制流缺陷

- 逻辑/结构死代码
- Switch语句中break遗失
- 非本地资源不当使用

### C/C++ 缺陷-Part 2

### 程序假死

- 死循环
- 双重锁或解锁丢失
- 负循环边界值
- 线程死锁
- 持锁过程中调用sleep()

### 空指针引用

- Null检查后引用空指针
- 直接引用返回的空指针
- Null检查前引用空指针

### 错误处理缺陷

- 未验证的返回值
- 未获取异常
- 负变量不当使用

### 代码维护性缺陷

- 多返回语句
- 无效变量

### 异常代码

- 复制/粘贴错误
- •Inte格式错误ftware Development Conference

### 不安全的数据处理

- 不可信的循环数据源
- 使用非可信数据源读写数组/指针
- 使用非可信数据源格式化字符串

#### 性能缺陷

- 值传递大参数
- 使用大堆栈

### 安全措施违反

- 缓冲区溢出
- 固定长度缓冲区写入
- 非安全函数调用
- 非安全临时文件使用
- 检查/使用时间不一致
- 用户空间指针不当使用

### API错误使用

- 非安全chroot调用
- 错误的迭代器使用
- printf()参数不匹配



### 安全检查

### • 定义Tainted Data

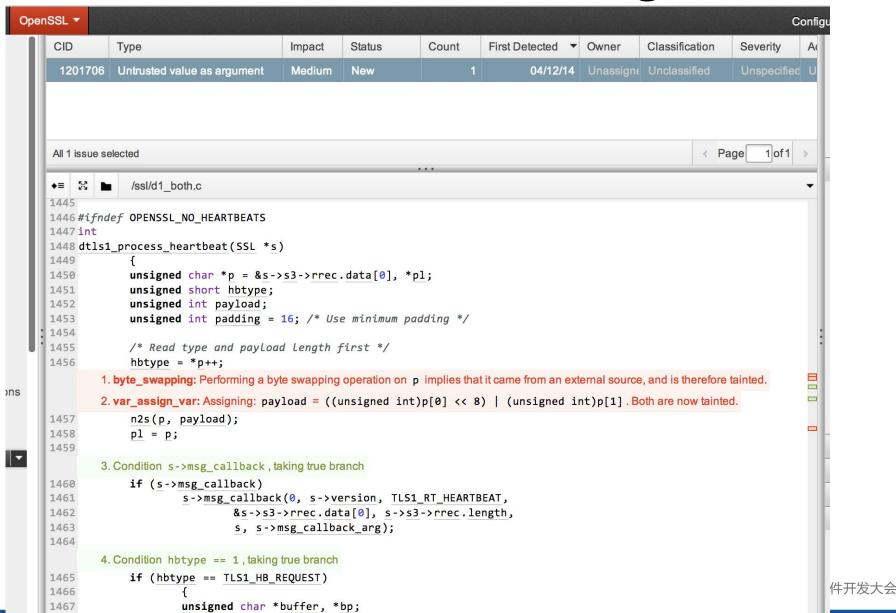
```
unsigned int packet_get_int() {
    return __coverity_tainted_data_return__();
}

void custom_read(int fd, void *buf) {
    __coverity_tainted_data_argument__(buf);
}
```

### HeartBleed

```
static unsigned int *__COV_TUI(unsigned int *i) {
    __coverity_tainted_data_argument__(i);
    return i;
}
#define TUI(i) *(__COV_TUI(&i))
#nodef n2s(c,s) n2s(c,TUI(s))
```

### 检测实例-HeartBleed Bug



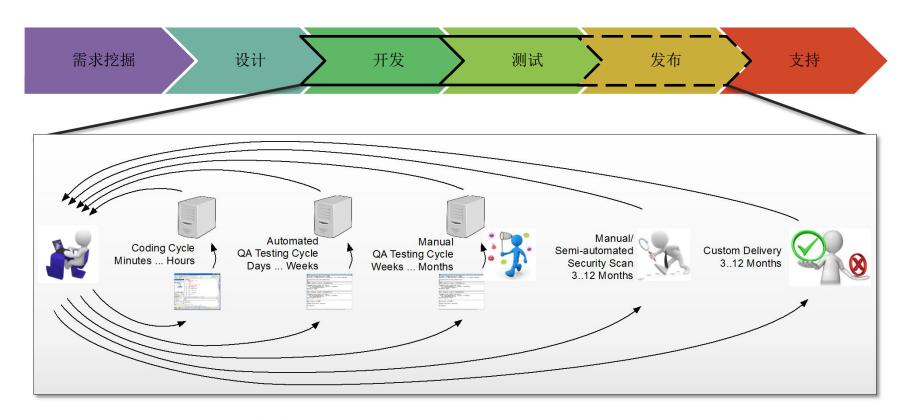
### Jenkins检测对比

类型	Coverity	FindBugs	Shared Defects
未处理的缺陷(Null引用)	79	7	5
资源泄露	86	12	13
并发问题	22	10	9
重要的缺陷	187	29	27
代码规范,最佳实践等	9	598	1
Bug 总数	196	627	28

### Freeradius缺陷检测对比

类型	Coverity	Clang	Shared Defects
内存问题	11	5	0
资源泄露	9	3	0
控制流缺陷,并发访问等 问题	83	30	1
重要的缺陷	103	38	1
代码规范,最佳实践等	9	59	2
Total Bugs	121	97	3

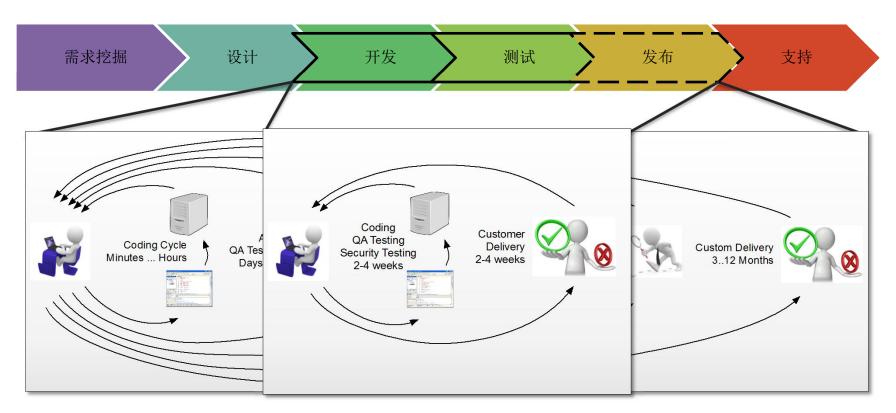
### "瀑布式开发流程"



### 经典的瀑布式开发流程:

- 代码开发与测试之间存在延迟
- QA经常由其他部门负责,在编码阶段完成后才能开始
- 安全保护基本是在"事发之后"才由单独的安全部门提供

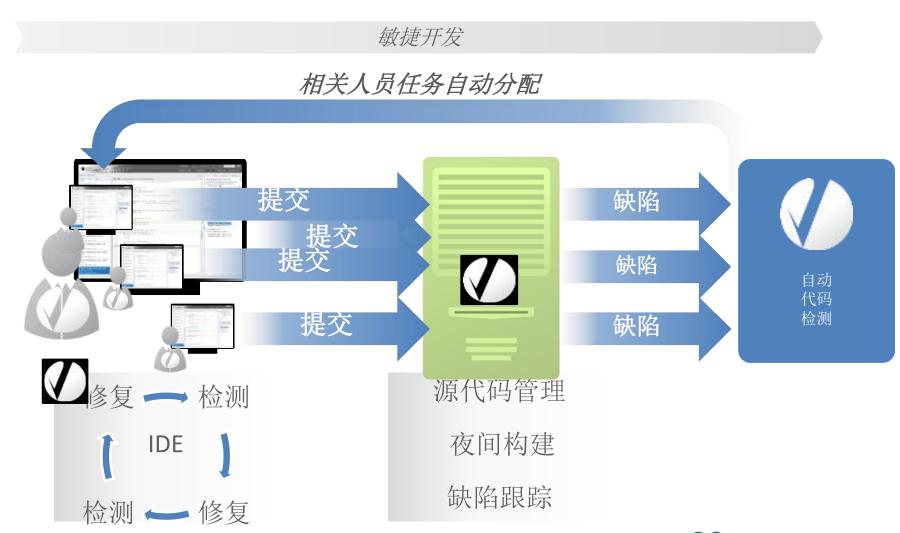
### "敏捷开发流程"



### 理想的敏捷开发流程:

- 将整个实施和验证流程缩短成2-4周的"冲刺"(sprints)
- 开发好的功能只有在经过全面验证之后才被"接受"
- 为取得成功, QA及安全测试都必须是深入的自动化测试

### 软件生命周期集成



### 示例: C/C++

- 这片代码来自PuTTY
  - PuTTY是面向Win32和Unix平台以及xterm终端仿真程序而免费实施的Telnet与SSH

```
static char *foo(char *s, int len)
    char *p = malloc((len + 1) * sizeof(char));
    memcpy(p, s, len);
    p[len] = '\0';
    return p;
                char *p = snewn(len + 1, char);
                memcpy(p, s, len);
                p[len] = '';
                return p;
```

### CVE-2005-0467

COMPATIBLE PRODUCTS

NEWS - FEBRUARY 27, 2012

SEARCH

#### Common Vulnerabilities and Exposures

Printer-Friendly View

The Standard for Information Security Vulnerability Names

TOTAL CVEs: 49296

HOME > CVE > CVE-2005-0467 (UNDER REVIEW)

#### **About CVE**

Terminology Documents FAOs

#### **CVE List**

About CVE Identifiers Search CVE Search NVD Updates & RSS Feeds Request a CVE-ID

#### CVE In Use

CVE Adoption CVE-Compatible Products NVD for CVE Fix Information

#### **News & Events**

Calendar Free Newsletter

#### Community

CVE Editorial Board Sponsor

#### Contact Us

Search the Site

#### CVE-ID

#### CVE-2005-0467 (under review)

CVE LIST

Learn more at National Vulnerability Database (NVD)
 Severity Rating • Fix Information • Vulnerable Software Versions • SCAP Mappings

#### Description

Multiple integer overflows in the (1) sftp\_pkt\_getstring and (2) fxp\_readdir\_recv functions in the PSFTP and PSCP clients for PuTTY 0.56, and possibly earlier versions, allow remote malicious web sites to execute arbitrary code via SFTP responses that corrupt the heap after insufficient memory has been allocated.

#### References

**Note:** <u>References</u> are provided for the convenience of the reader to help distinguish between vulnerabilities. The list is not intended to be complete.

- IDEFENSE:20050221 Multiple PuTTY SFTP Client Packet Parsing Integer Overflow Vulnerabilities
- URL:http://www.idefense.com/application/poi/display?id=201&type=vulnerabilities
- CONFIRM:http://www.chiark.greenend.org.uk/~sqtatham/putty/wishlist/vuln-sftp-string.html
- CONFIRM:http://www.chiark.greenend.org.uk/~sgtatham/putty/wishlist/vuln-sftp-readdir.html
- CONFIRM:http://www-1.ibm.com/support/docview.wss?uid=ssq1S1002414
- CONFIRM:http://www-1.ibm.com/support/docview.wss?uid=ssq1S1002416
- GENTOO:GLSA-200502-28
- URL:http://www.gentoo.org/security/en/glsa/glsa-200502-28.xml
- SECUNIA:14333
- URL:http://secunia.com/advisories/14333
- SECUNIA:17214
- URL:http://secunia.com/advisories/17214
- XF:putty-sftppktgetstring-bo(19403)
- URL:http://xforce.iss.net/xforce/xfdb/19403

#### **CVE List**

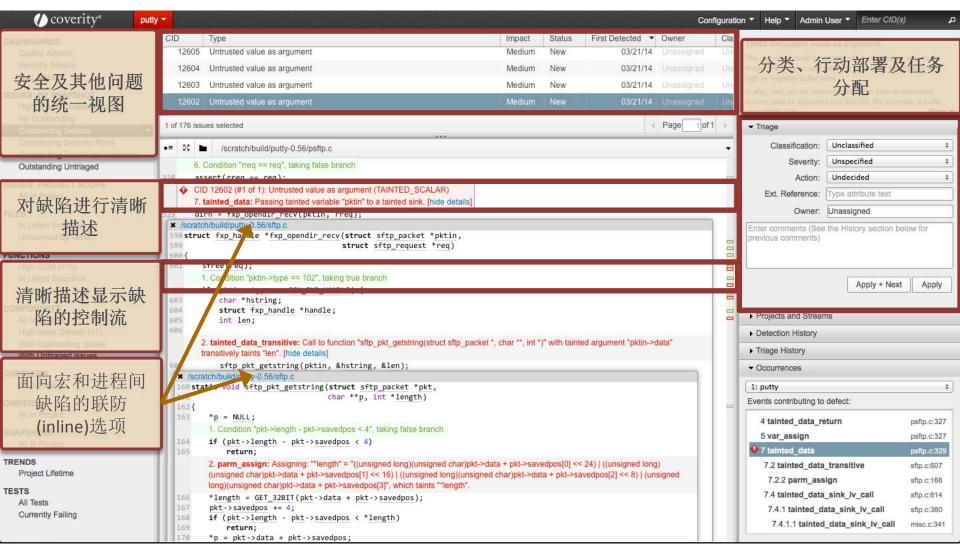
About CVE Identifiers
Editorial Policies
Data Sources
Reference Key/Maps
Search Tips
Updates & RSS Feeds
Obtain a CVE Identifier

#### ITEMS OF INTEREST

Terminology NVD

```
static void sftp pkt getstring(struct sftp packet *pkt,
                                char **p, int (*length))
    *p = NULL;
    if (pkt->length - pkt->savedpos <
        return:
    /* length value is taken from aser-supplied data
    *length = GET 32BIT (pkt->data + pkt->savedpos);
    pkt >savedpos += 4;
    /* this check will be wassed if length < 0 */
    if (pkt->length - pkt->savedpos < *length)</pre>
       return;
    *p = pkt->data + pkt >savedpos;
   pkt->savedpos += *length;
        /* sftp pkt getstring call with control 16d len value */
        sftp_pkt_getstring(pktin, &hstring &len);
        handle = snew(struct fxp handle);
        /* heap corruption will occur if len == -1 */
        handle->hstring = mkstr(hstring(len))
        handle->hlen = len;
                                    static char *mkstr(char *s,/int len)
        sftp pkt free (pktin);
        return handle;
                                        char *p = snewn(len + 1, char);
                                        memcpy(p, s, len);
                                        p[len] = '';
                                        return p;
```

### 如何处理?







10号展位◎





## THANKS!