开源软件更可信更安全？
----从开源代码安全到开源生态安全

王怀民

国防科学技术大学
National University of Defense Technology

内容

问题的本质

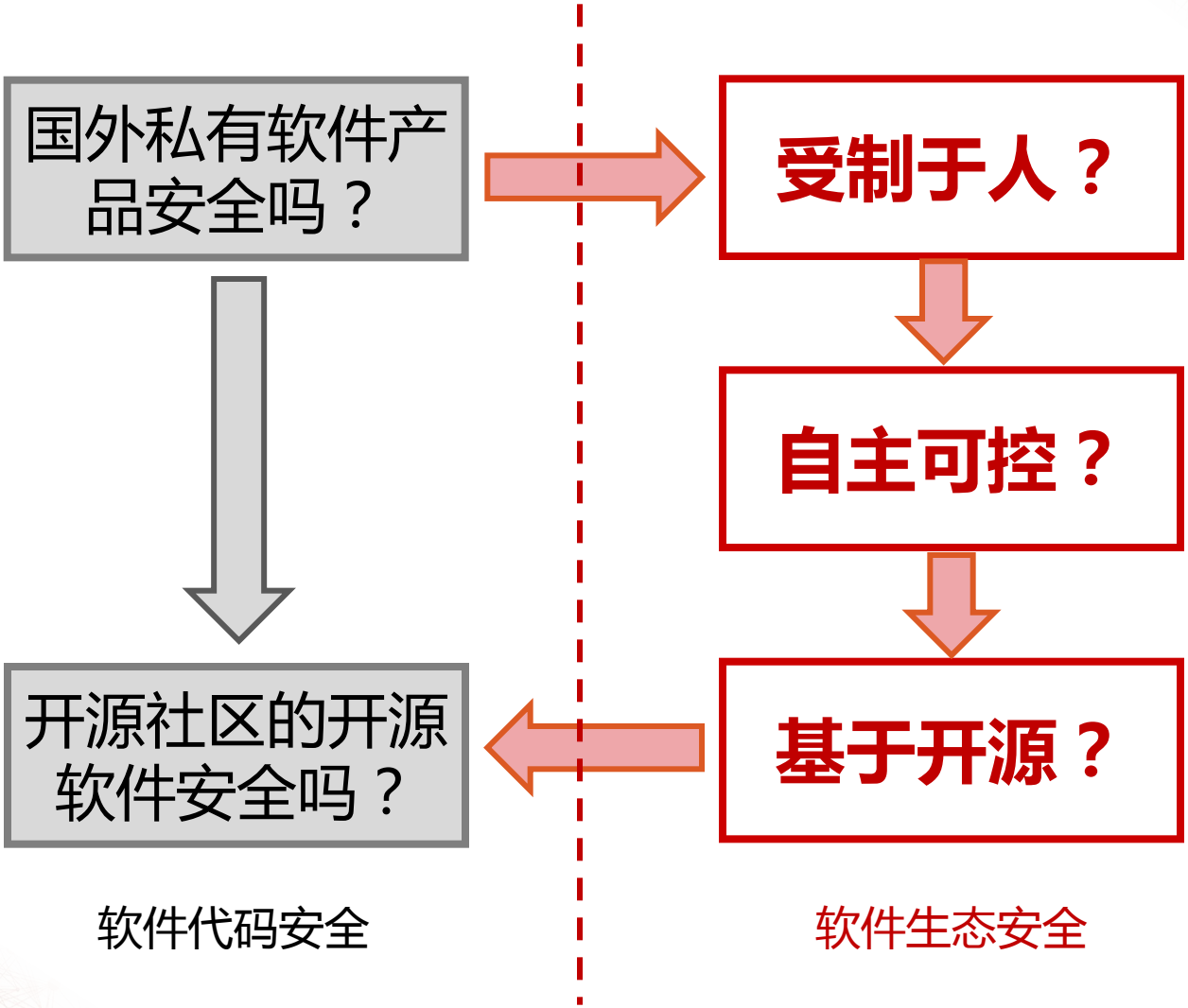
努力的方向

内容

问题的本质

努力的方向

问题的演变



代码质量问题

开源与私有C/C++项目代码质量对比

	开源代码	私有代码
代码总行数	252,010,313	684,318,640
项目数	741	493
平均项目规模（代码行数）	340,094	1,388,070
截止2013年底重大缺陷个数	149,597	492,578
2013年修复的缺陷数量	44,641	783,799
缺陷密度	**0.59**	**0.72**

摘自《2013-Coverity-Scan-Report》

代码质量问题

不同规模的开源与私有C/C++项目缺陷密度对比

项目代码规模（代码行）	开源代码	私有代码
<100,000	0.35	0.38
100,000-499,999	0.50	0.81
500,000-1,000,000	0.70	0.84
>1,000,000	0.65	0.71
平均缺陷密度	**0.59**	**0.72**

摘自《2013-Coverity-Scan-Report》

优秀的开源软件代码质量不逊于优秀的私有软件代码

代码质量问题

不可能证明软件代码没有漏洞
不可能证明两个软件代码的质量谁更优

OpenSSL
Cryptography and SSL/TLS Toolkit

The Heartbleed Bug

The Heartbleed Bug is a serious vulnerability in the popular OpenSSL cryptographic software library. This weakness allows stealing the information protected, under normal conditions, by the SSL/TLS encryption used to secure the Internet. SSL/TLS provides communication security and privacy over the Internet for applications such as web, email, instant messaging (IM) and some virtual private networks (VPNs).

The Heartbleed bug allows anyone on the Internet to read the memory of the systems protected by the vulnerable versions of the OpenSSL software. This compromises the secret keys used to identify the service providers and to encrypt the traffic, the names and passwords of the users and the actual content. This allows attackers to eavesdrop on communications, steal data directly from the services and users and to impersonate services and users.

What leaks in practice?

We have tested some of our own services from attacker's perspective. We attacked ourselves from outside, without leaving a trace. Without using any privileged information or credentials we were able steal from ourselves the secret keys used for our X.509 certificates, user names and passwords, instant messages, emails and business critical documents communication.

How to stop the leak?

As long as the vulnerable version of OpenSSL is in use it can be abused. Fixed OpenSSL has been released and now it has to be deployed. Operating system vendors and distribution, appliance vendors, independent software vendors have to adopt the fix and notify their users. Service providers and users

Errata Security

Advanced persistent cybersecurity

Wednesday, September 24, 2014

Bash bug as big as Heartbleed

By Robert Graham

Today's *bash* bug is as big a deal as Heartbleed. That's for many reasons.

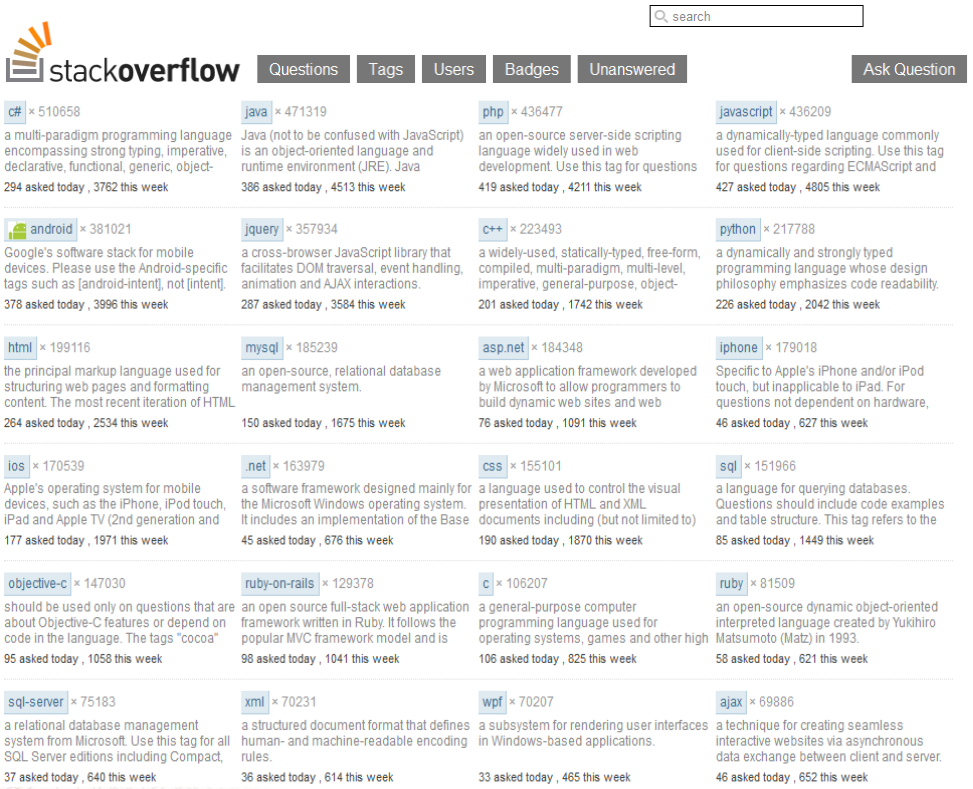
The first reason is that the bug interacts with other software in unexpected ways. We know that interacting with the shell is dangerous, but we write code that does it anyway. An enormous percentage of software interacts with the shell in some fashion.

ll the software out there that is vulnerable to
SSL bug: OpenSSL is included in a bajillion
ble to fully quantify exactly how much

阳光下软件更安全？

开发成本问题

降低获得和使用可信证据的成本
降低软件供应链安全管理的风险和成本



整体情况

注册用户：	>190万
发布问题：	>714万
获得回答：	>1251万
问题回答比例：	>92%
首次响应：	<11分钟

以Android为例

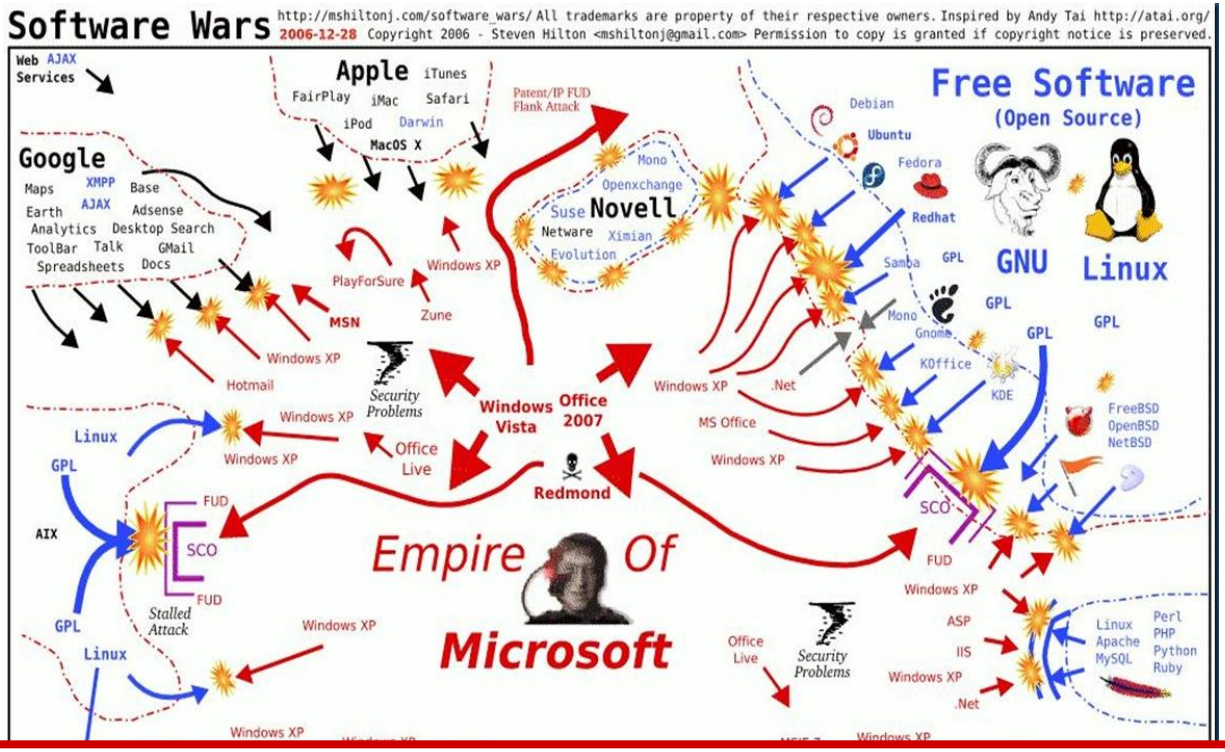
参与讨论用户：	>3.5万
覆盖API：	>87%
浏览次数：	>7000万

开发成本问题

降低复用高质量软件的成本

- Linux操作系统包含1700万行代码，MySQL数据库有1269万行代码，Apache应用服务器有227万行代码，Google Chrome浏览器有770万行代码
- 2013年，全球80％的新出现软件产品含有开源软件代码， 98％的全球化企业使用开源软件
- 福布斯排名前两千的企业使用的软件，平均30％的代码来自开源软件，最高达80％

发展模式问题



开源软件与私有软件之战
是保护创新与保护创收的博弈
是软件开发模式及其生态环境的博弈

发展模式问题

产品：面向用户需求

生产：质量目标控制

问题的本质

代码生成问题	→	代码在创新生态环境中 持续成长演化问题
代码安全问题	→	代码生态环境的安全问题
代码可信问题	→	帮助使用者理解和掌控代码 及其可信证据链的问题

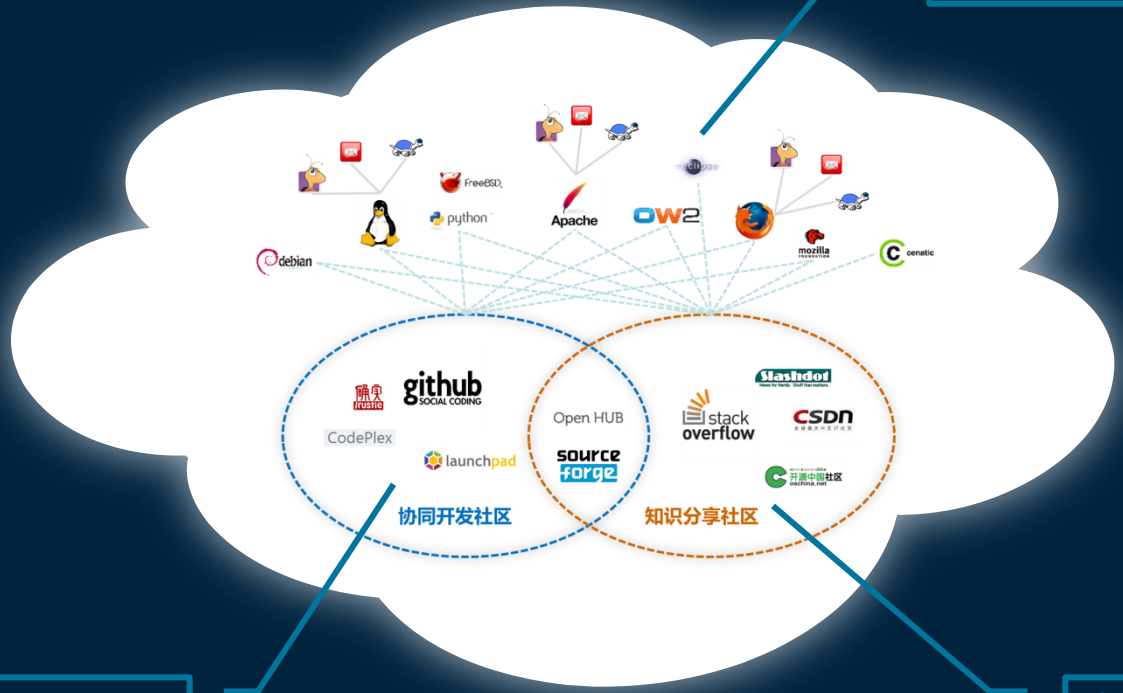
内容

问题的本质

努力的方向

深入了解开源社区

有多少开源组织？

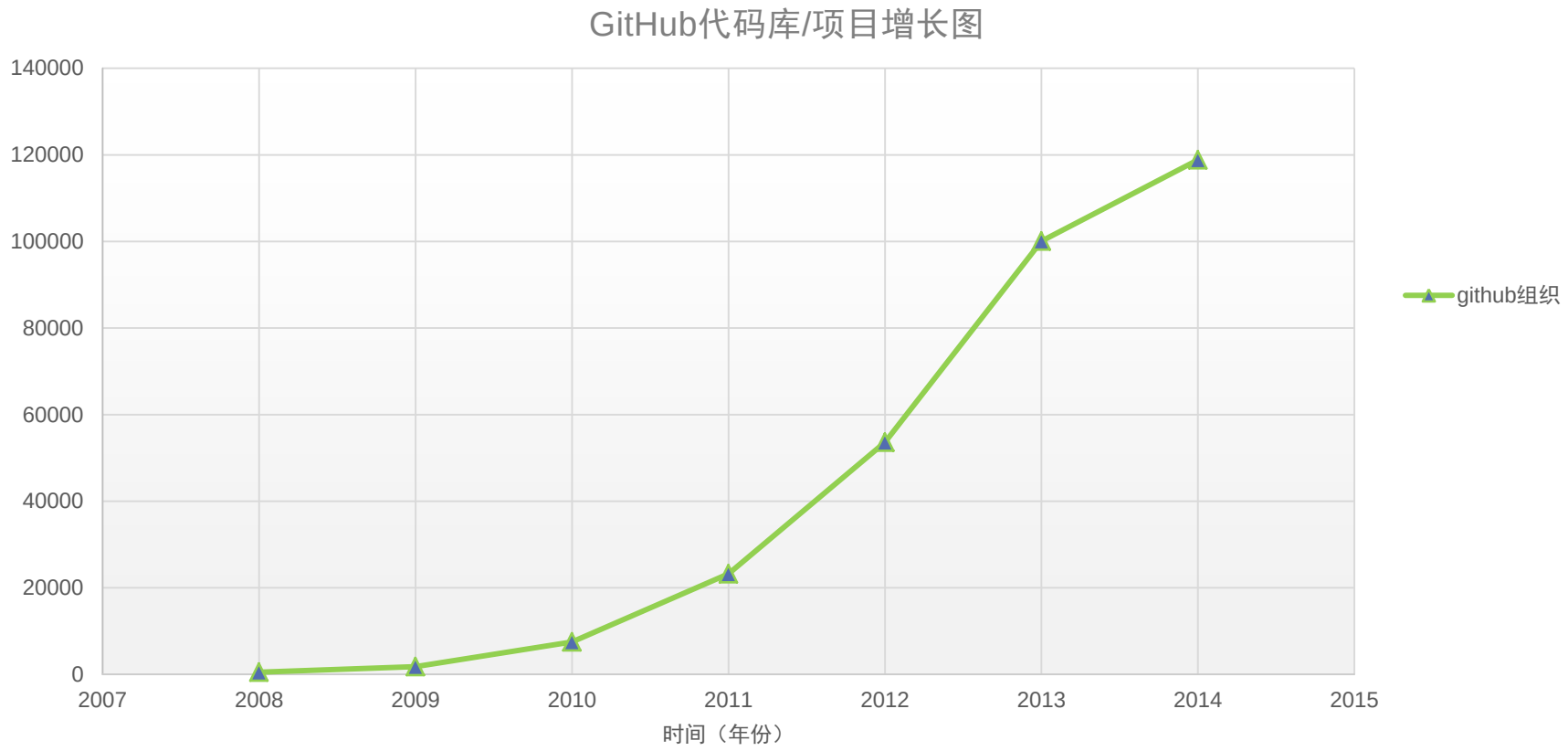


有多少开源项目？

有多少开源消息？

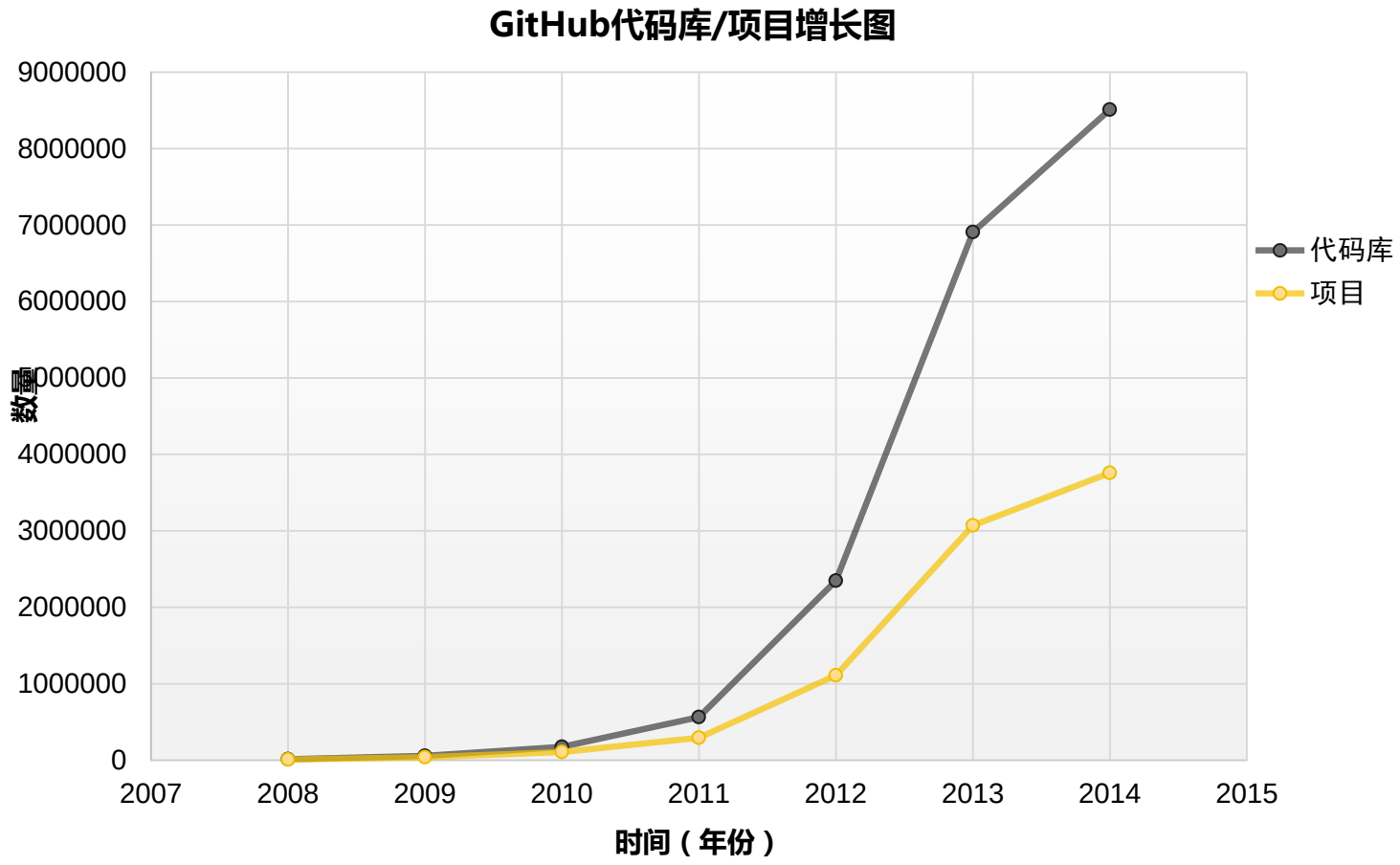
开源组织的增长

✓ 截至2014年4月，GitHub中的开源组织超过11.8万个



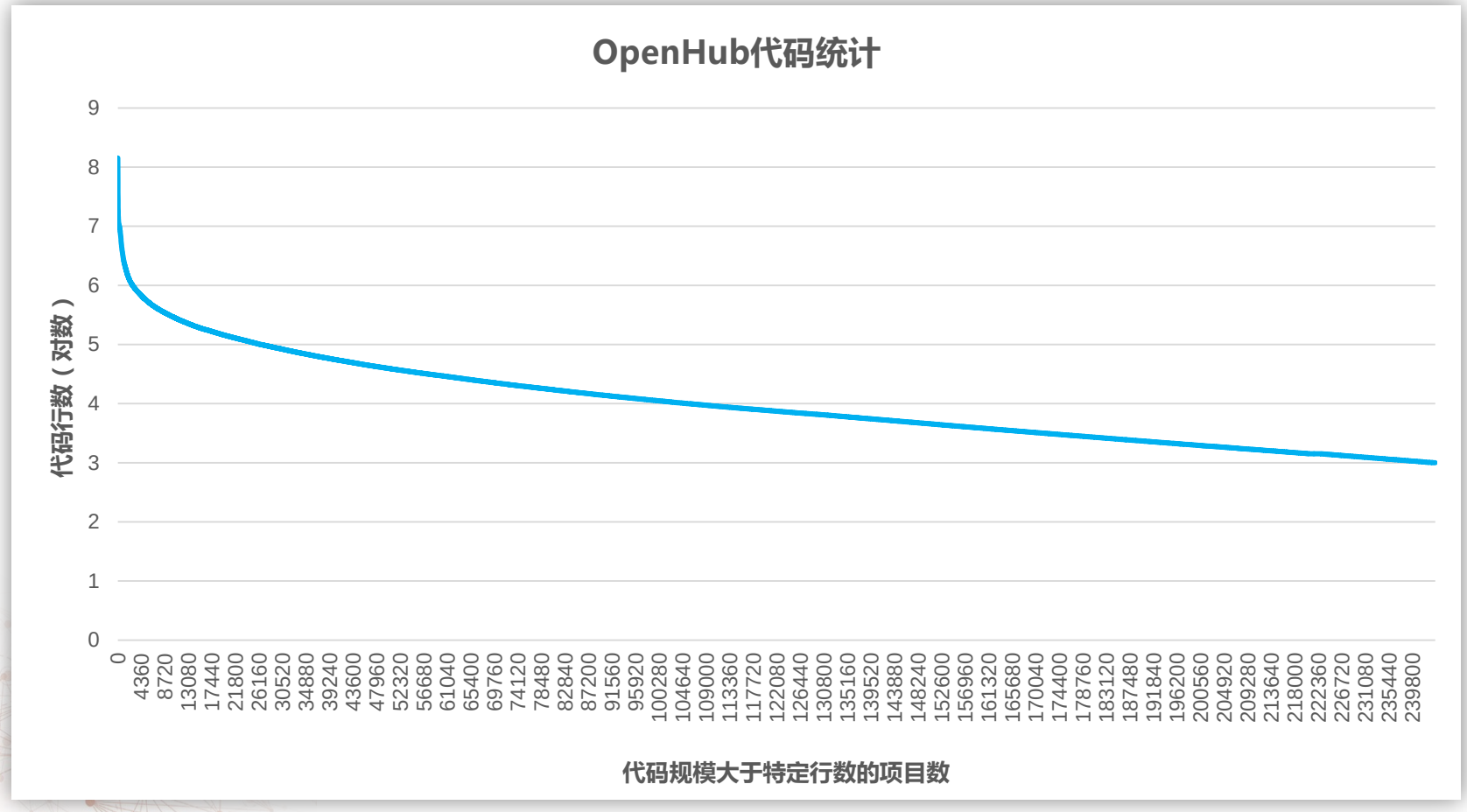
开源项目的增长

✓ 截至2014年4月，GitHub中的代码库数量超过850万，项目数量超过375万



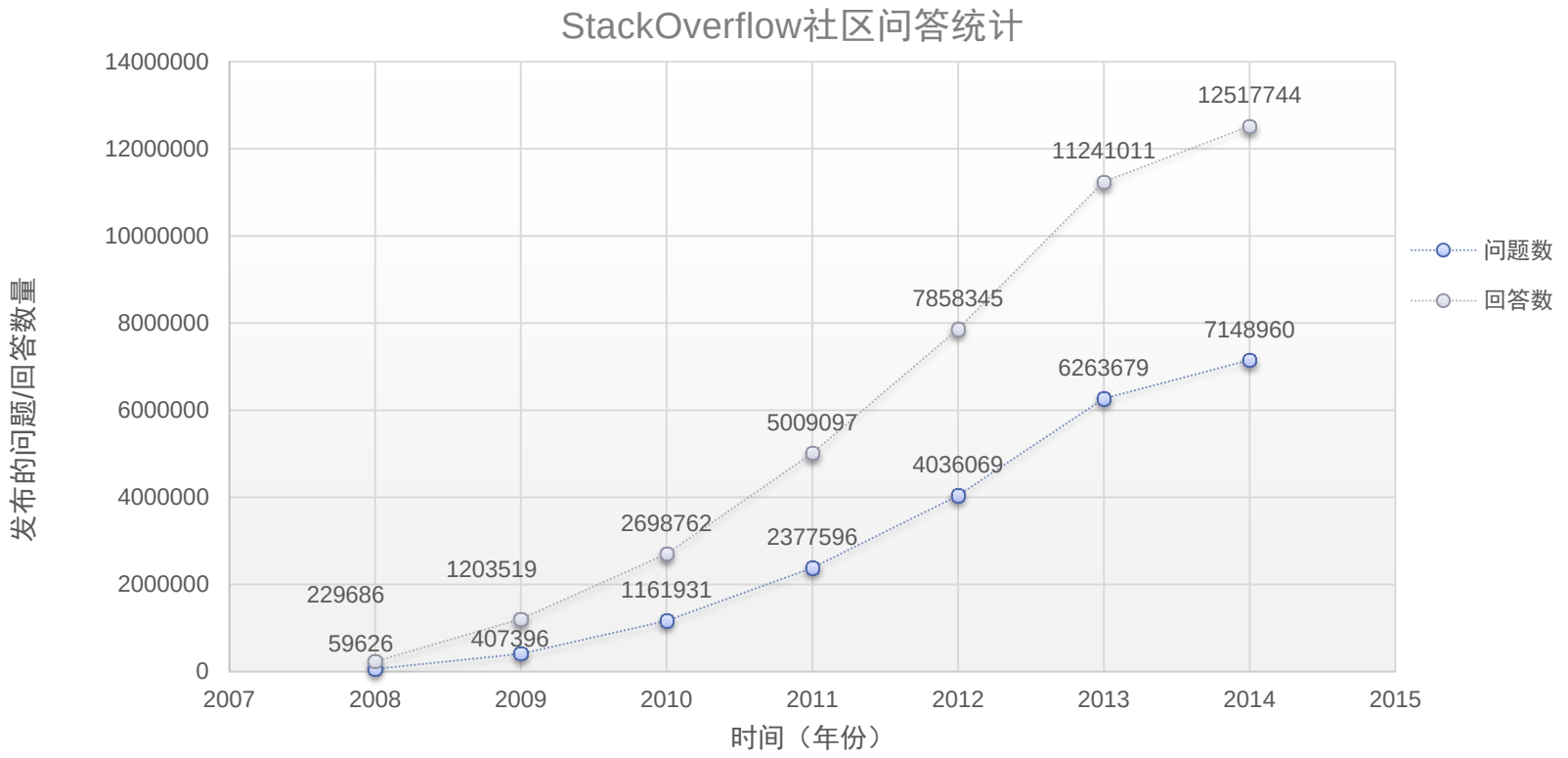
开源代码的规模

- ✓ **截至2014年8月，OpenHub中开源软件代码总规模超过217亿**
  - 代码行数超过1万行的项目超过10万个
  - 代码行数超过1070万的项目为200个



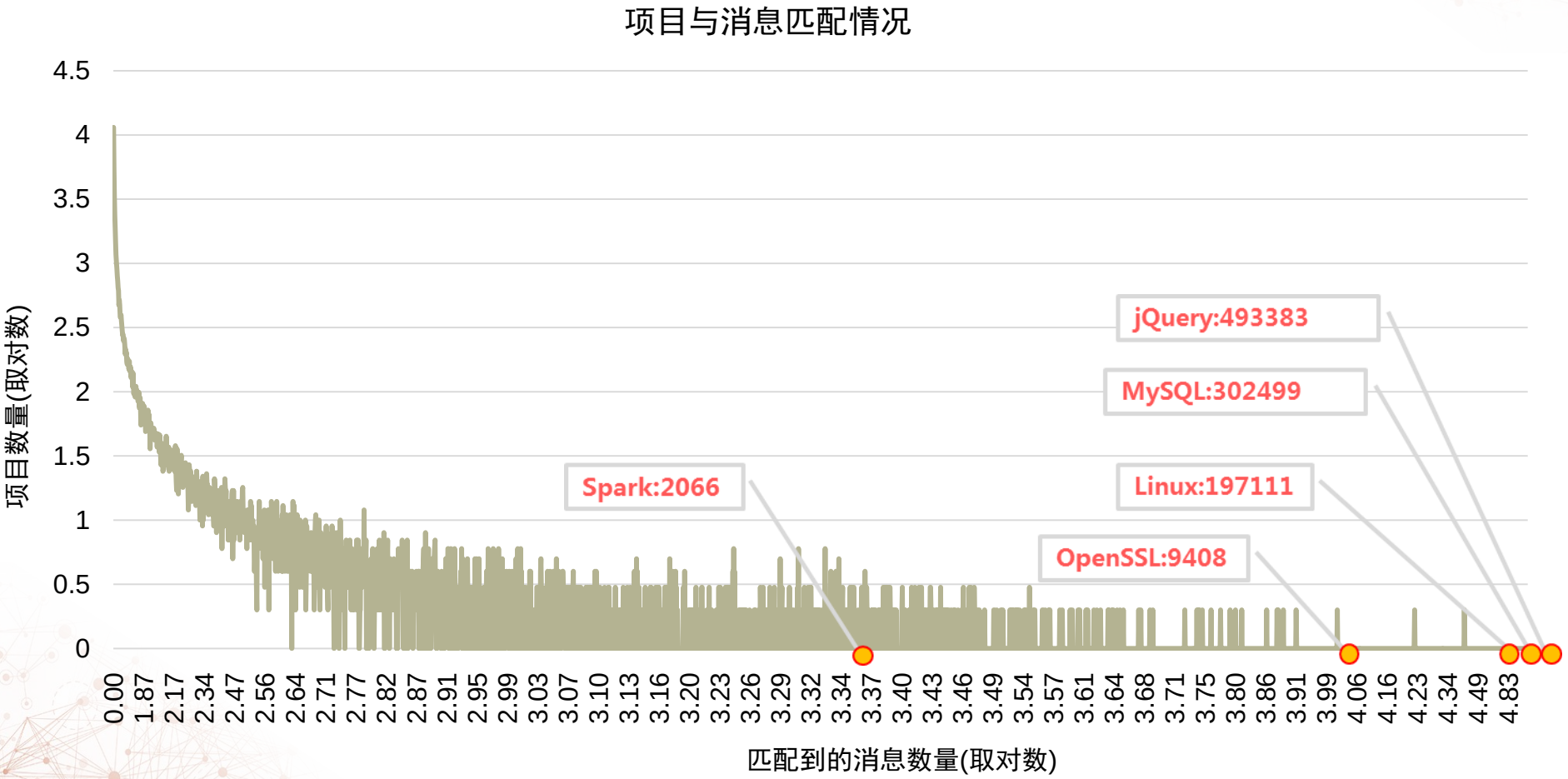
开源项目的问答

- ✓ **截至2014年5月，StackOverflow问题数量超过710万，回答数超过1200万，且保持快速增长**
  - 总人数超过190万，普通浏览用户更多
  - 参与帖子数超过5个的人员占24%



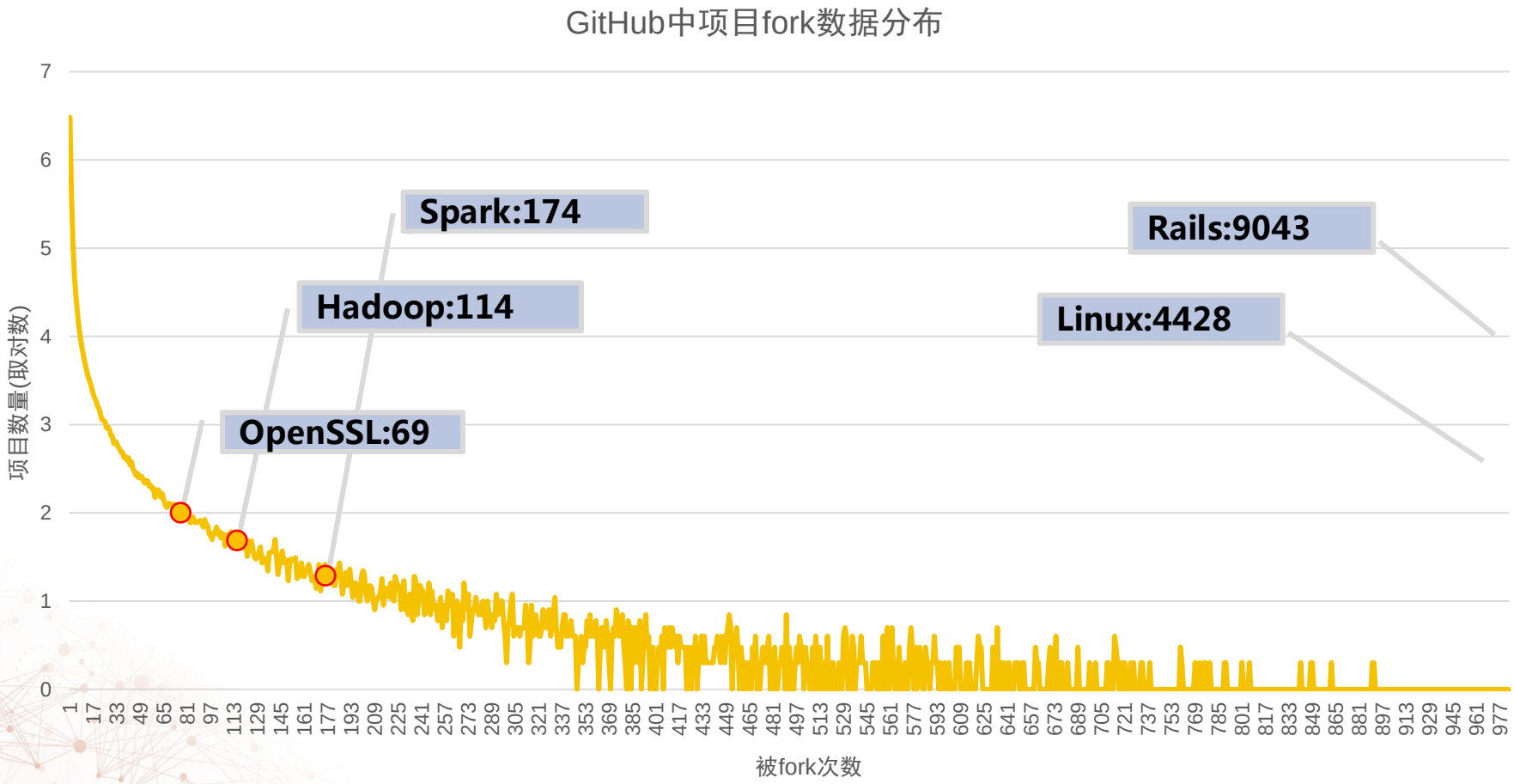
开源项目的评论

- ✓ **在27万个开源项目中，有超过5万个项目在知识分享社区中被讨论，讨论的频度呈现出长尾特征**

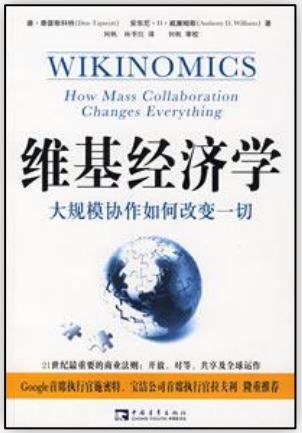


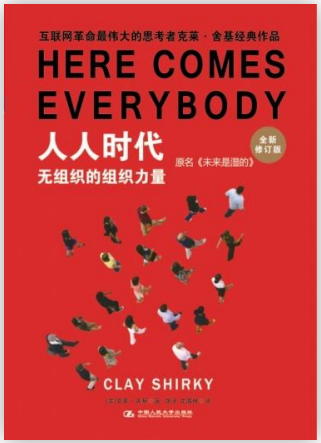
开源项目的演化

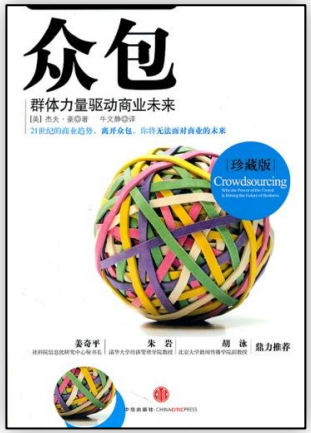
- ✓ 截至2014年4月，GitHub中fork次数>10次的项目数约为5.4万，约占项目总数的1.5%

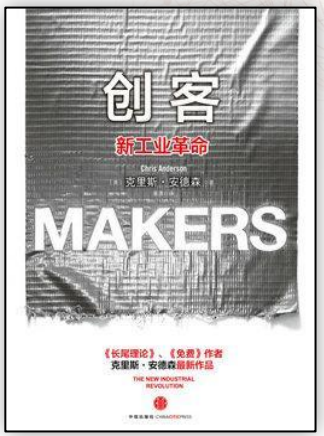


深入认识开源机制









自由对等的互联网创新实践

互联网创作实践的两大核心机理

- 群体协同：大规模对等协同模式催生出大量创新作品
- 持续演化：作品在网络环境获得持续反馈和不断改进

互联网协同创作机理 → **系统的引入…** → 基于群体智慧的群体化软件开发方法

成长演化的软件观

成长性构造
法则

好的开源软件是在持续积累的过程中发展而成的

适应性演化
法则

好的开源软件是在不断适应环境和需求的变化过程中持续演化的

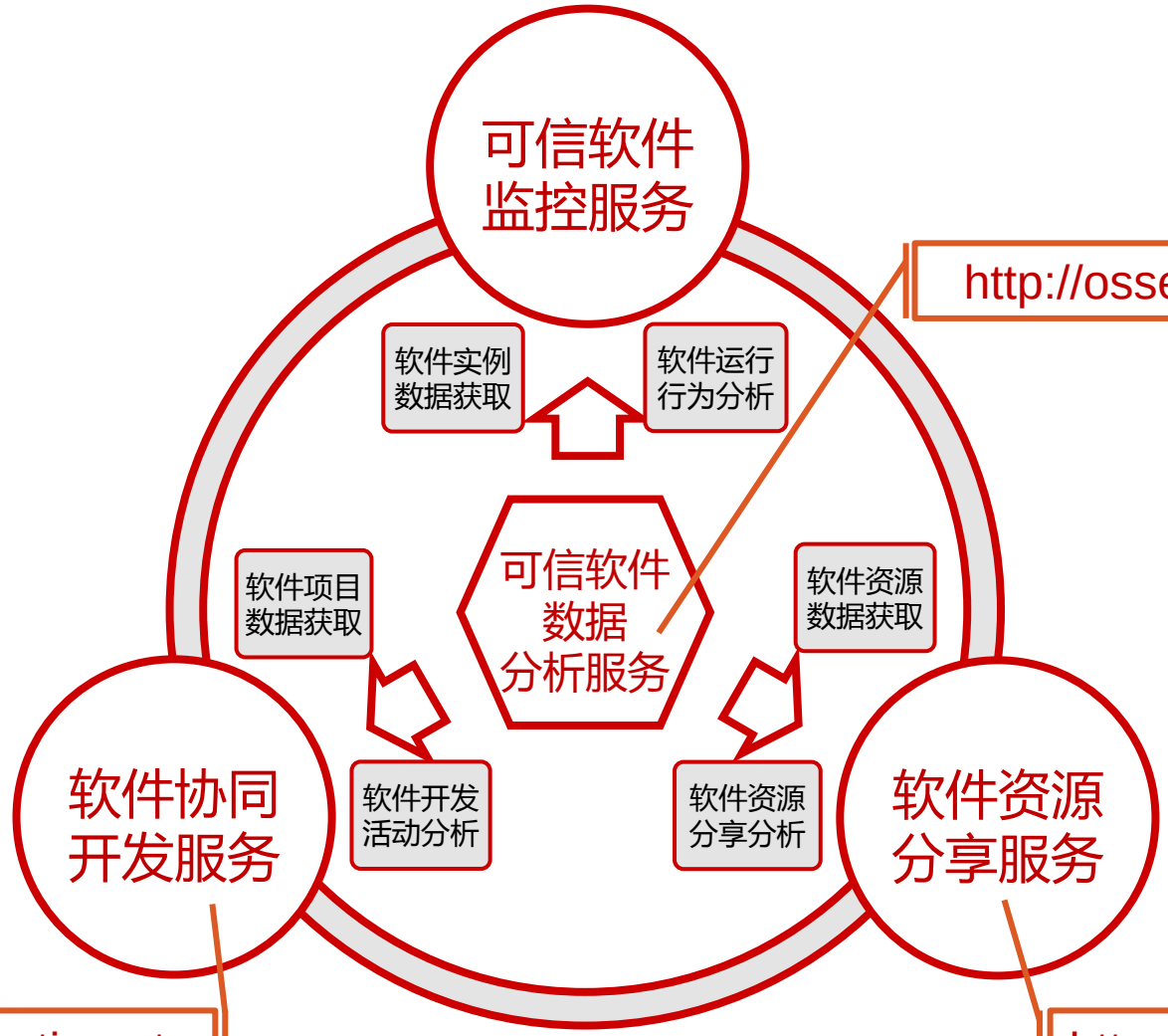
群体化的软件开发方法

代码与证据紧密耦合的
可信软件演化模型

创作与生产紧密耦合的
大规模协同开发过程模型

协同、共享、监控与分析
紧密耦合的服务支撑模型

群体化的软件开发服务平台



结论

无法证明开源代码更安全

可以相信开源生态更安全

需要关注群体化方法研究

谢谢！问题？