

中国Linux内核开发者大会

创新驱动开源：华为在Linux社区的历程与思考

陈海波 华为OS内核实验室



openEuler公众号





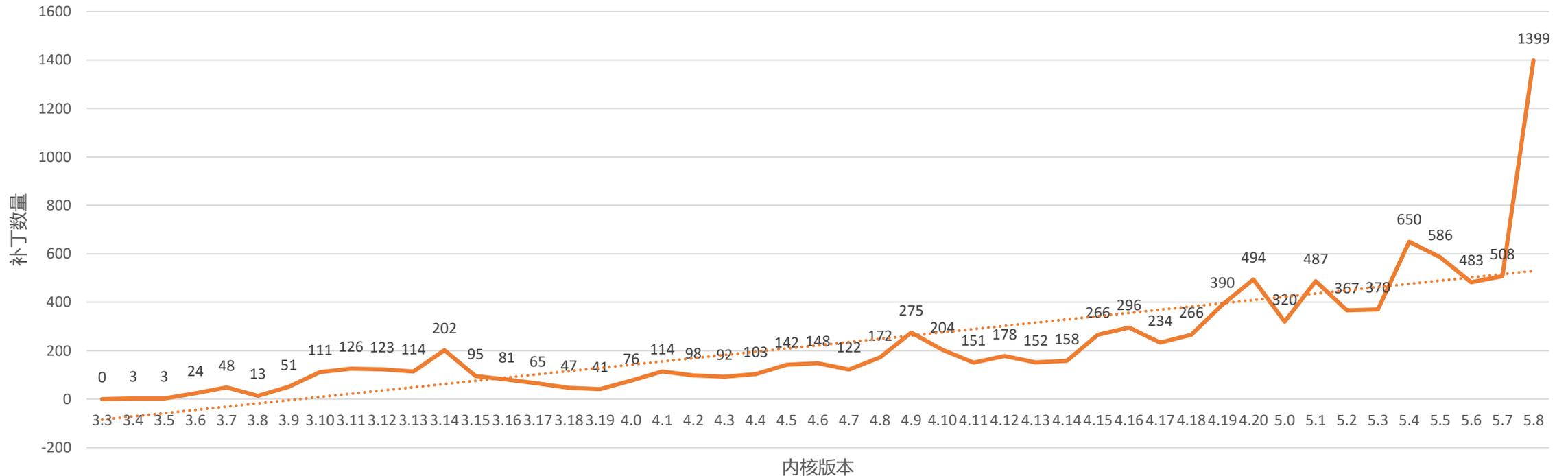
- 1 / 华为在Linux内核的贡献历程
- 2 / 业务创新驱动开源
- 3 / 聚焦的方向和思考
- 4 / 小结

华为在Linux内核社区的贡献历程

华为从2012年开始贡献Linux内核，社区贡献稳步上升。

- 累计超过**11000**个补丁被主线接受；
- 在软硬件解耦ACPI on ARM64，文件系统F2FS/EROFS，RAS EDAC框架，Media子系统，IIO 子系统，容器所使用的核心功能 Cgroup，以及 Perf on ARM64 等多个子系统上，华为**担任社区Maintainer**席位。

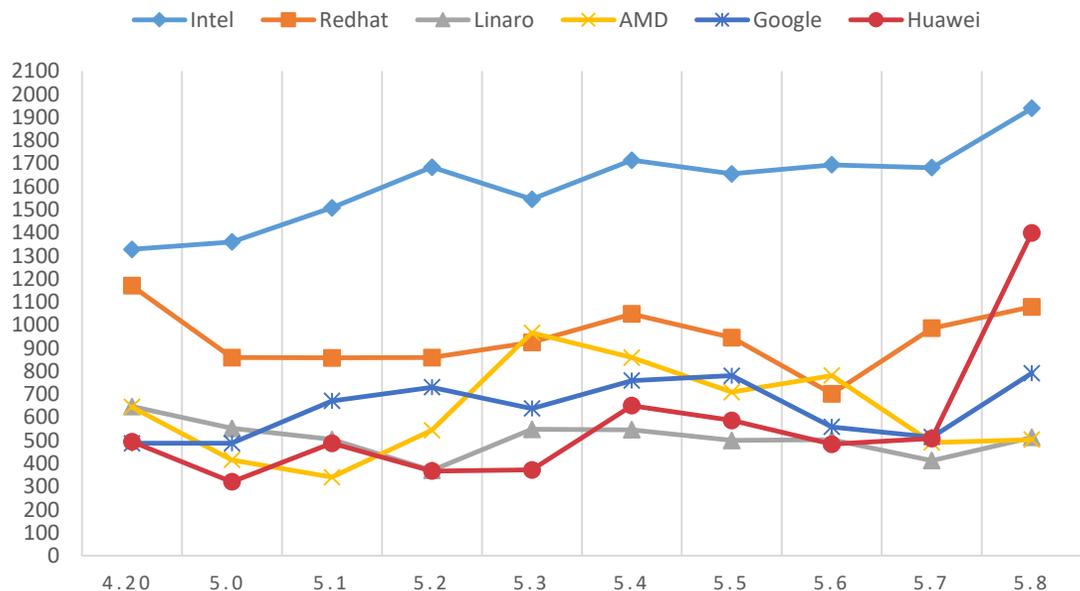
华为各内核版本贡献的补丁数量



5.8内核，Linux内核社区补丁贡献TOP2

从4.20内核版本开始，华为贡献持续在TOP10以内

LINUX KERNEL 厂商代码贡献



For the most part, this table looks fairly familiar, but the fact that Huawei has moved up to the top of the list may come as a bit of a surprise. Much of this is the result of Chehab's work described above, but Huawei's contribution this time around is rather larger than that. A great deal of effort has gone into freezing Huawei out of the commercial marketplace in significant parts of the world, but the company remains active in the development community with 92 developers contributing to 5.8. For the curious, Huawei's work was mostly focused in these subsystems:

5.8内核版本，华为贡献补丁数量排名TOP2

- 贡献补丁1399个
- 代码修改行数排名第一
- HULK Robot发现缺陷数量排名第一

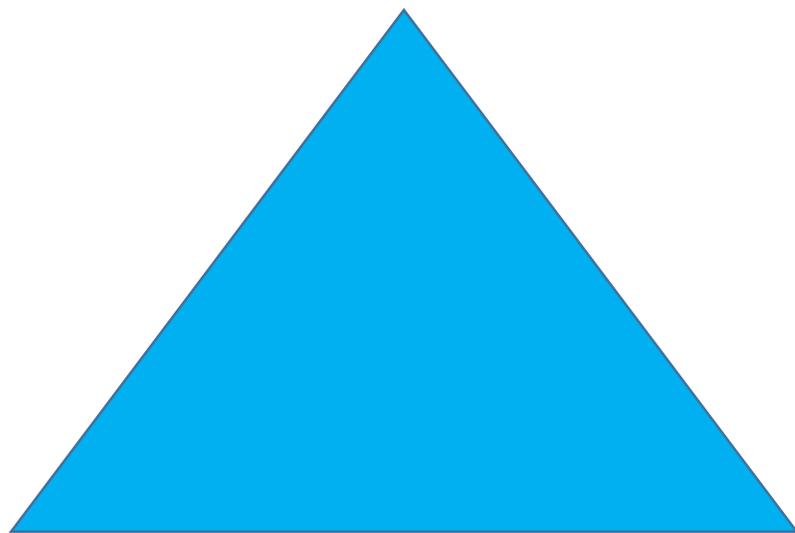
Most active 5.8 employers					
By changesets			By lines changed		
Intel	1939	11.9%	Huawei Technologies	293365	27.8%
Huawei Technologies	1399	8.6%	Habana Labs	93213	8.8%
(Unknown)	1231	7.5%	Intel	88288	8.4%
Red Hat	1079	6.6%	(None)	47655	4.5%
(None)	1016	6.2%	(Unknown)	36786	3.5%
Google	791	4.9%	Linaro	36322	3.4%
IBM	542	3.3%	Red Hat	34737	3.3%
(Consultant)	515	3.2%	Google	34209	3.2%
Linaro	513	3.1%	IBM	24233	2.3%
AMD	503	3.1%	Mellanox	23364	2.2%
SUSE	463	2.8%	Realtek	22767	2.2%
Mellanox	445	2.7%	AMD	21411	2.0%
NXP Semiconductors	330	2.0%	NXP Semiconductors	21328	2.0%
Renesas Electronics	322	2.0%	(Consultant)	15418	1.5%



- 1 / 华为在Linux内核的贡献历程
- 2 / 业务创新驱动开源
- 3 / 聚焦的方向和思考
- 4 / 小结

华为在Linux社区的驱动力：应用+硬件+技术

用户体验驱动的技术创新：
为上层应用提供更快、更安全、更易用的执行环境？
为底层硬件建立高效、安全、高利用率的抽象？



端管云全场景应用：
5G、云与计算、终端

全场景芯片与应用：
麒麟、昇腾、巴龙等

业务创新驱动开源贡献



- ARM64架构支持
- 软硬件解耦ACPI
- 内核热补丁
- 扩展性增强



- F2FS
- EROFS
- 分布式文件系统
- 智能调度



目录

- 1 / 华为在Linux内核的贡献历程
- 2 / 业务创新驱动开源
- 3 / 聚焦的方向和思考
- 4 / 小结

ICT领域 - 算力多样性支持

关键技术点

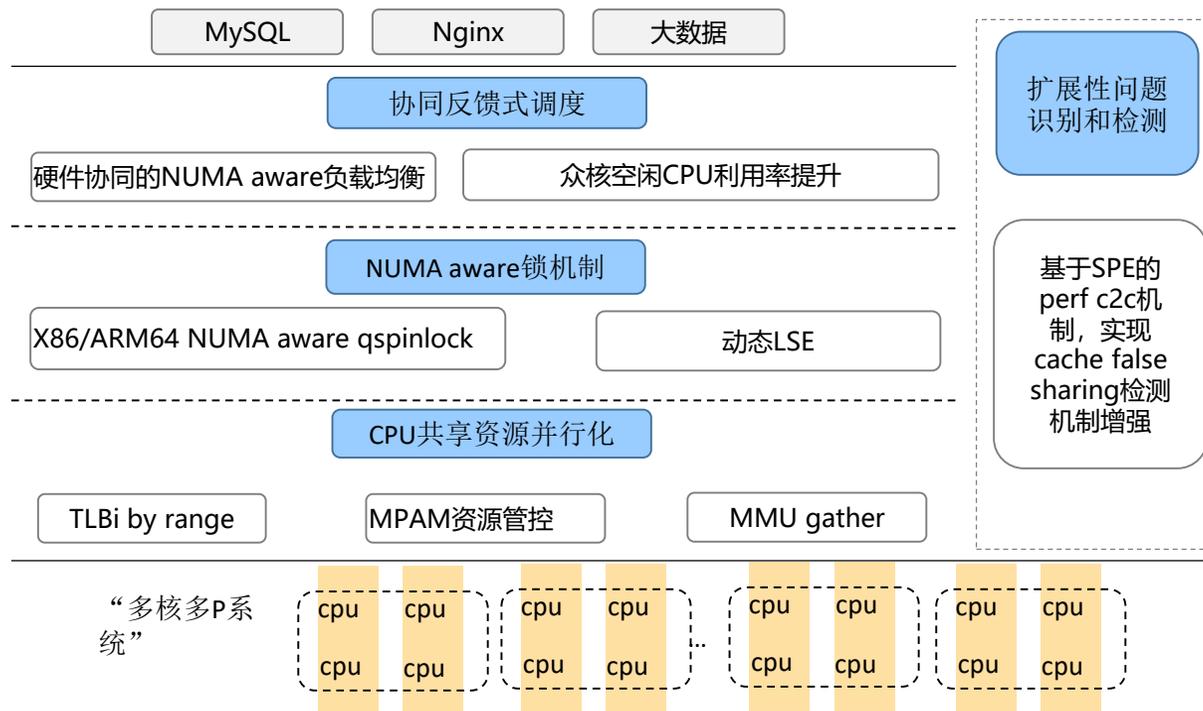
- ① **系统扩展性**: 服务器的多P多核系统, 通过调度、锁机制以及减少CPU共享资源冲突提升系统扩展性。
- ② **软硬件解耦**: 通过ACPI/UEFI/DT等软硬件解耦技术, 多CPU架构复用驱动代码, 降低OS与底层硬件耦合。
- ③ **CPU/XPU统一共享虚拟内存池**: SVA等技术实现CPU和XPU内存‘零’拷贝, 提供高性能底座。
- ④ **内核热补丁**: 运行时动态修复系统bug。
- ⑤ **HULK robot**: 内核稳定性保障。



通过调度、NUMA aware锁机制等提升系统扩展性

关键技术

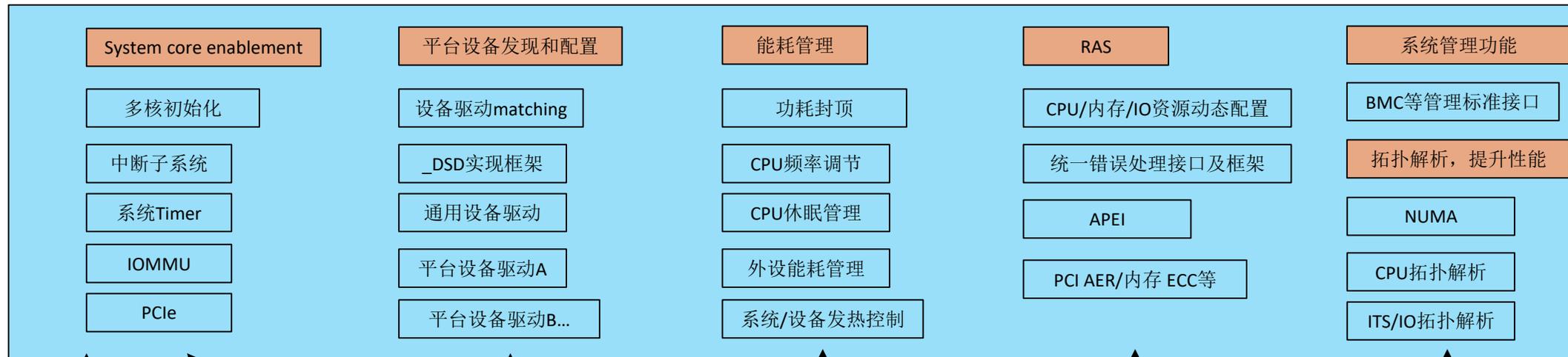
- 协同反馈式调度。通过软硬件协同，实现NUMA aware的精准负载迁移和均衡。
- NUMA aware锁机制。Qspinlock NUMA aware，NUMA域内优先获取锁，降低cache乒乓，提升系统吞吐量
- 减少CPU共享资源冲突。通过架构优化以及MPAM技术，降低TLB，cache和内存带宽的冲突，提升系统扩展性
- 提供扩展性问题识别和检测工具，识别app的扩展性问题（如cache false sharing）



通过ACPI架构解耦，实现多平台OS共享BIOS与驱动架构

ACPI user space daemon
(acpidaemon, rasdaemon)

User Space



Kernel



UEFI

ARM Trusted Firmware

Intel X86

Huawei 鲲鹏

AMD x86

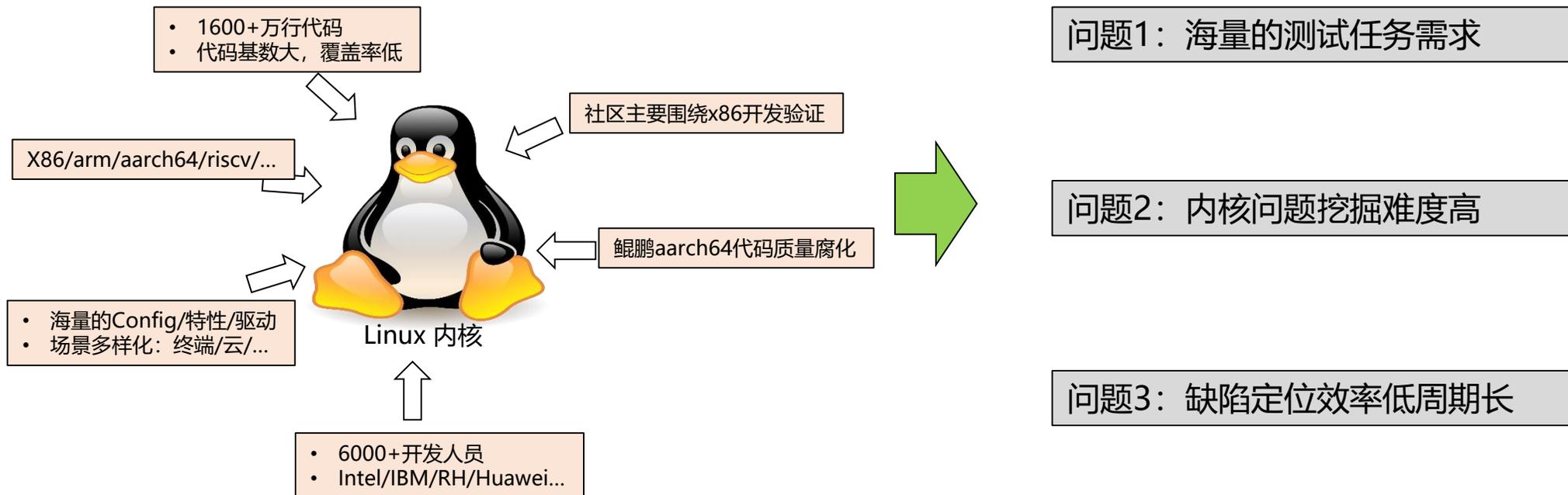
飞腾

海光

Hardware

HULK Robot - Linux开源内核的测试挑战

- **问题挑战:** 随着业务发展, 越来越多的业务使用开源组件构建版本或开源自身代码。开源模式下除了带来业务生态快速催熟等各种红利外, 也引入了越来越多的挑战: **海量频繁的补丁合入, 成千上万的开发人员, 一行修改百倍测试等。**



■ 方法思路:

- **分布式/弹性可扩展测试平台:** 分布式, 测试JOB管道化, 精准测试
- **缺陷挖掘:** 全系统的函数级故障注入技术, 基于语义的精准fuzz用例自动生成
- **缺陷定位/Fix:** 复现用例自动生成, 静态检查问题自动生成fix补丁, 传统二分, 缺陷自动定位

HULK Robot - 提升多场景内核稳定性，多次居Linux缺陷挖掘榜首，为社区贡献提供强力输入

开源社区质量保证流程融入Robot

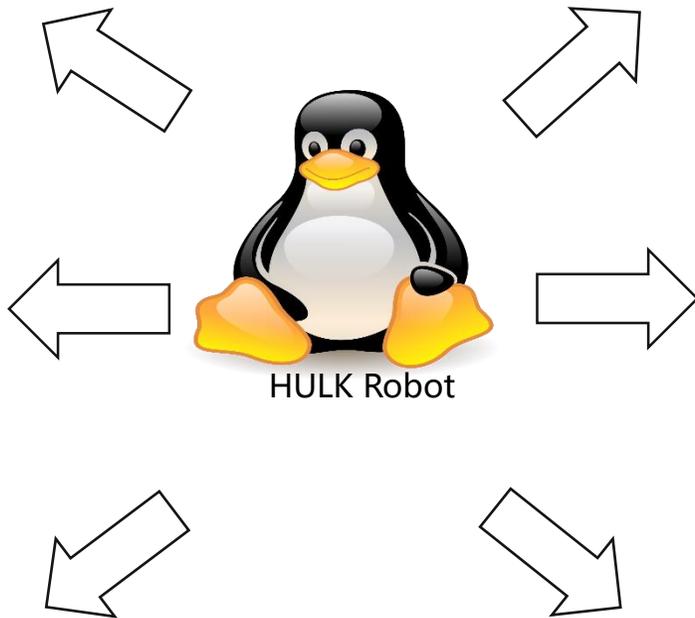


全体系结构覆盖

- Aarch64
- Arm
- X86
- RISCv...

全质量防护网

- 特性树功能全覆盖，囊括社区和
- 性能
- 兼容性
- 安全
- 长稳
- Fuzz
- ...



补丁级测试

- 补丁级代码全生命周期管理
- 补丁级精准测试

社区认可的问题持续挖掘技术

- 语义FUZZ技术，持续自动的问题和CVE挖掘
- 代码级精准测试覆盖，高效验证海量补丁
- 函数级故障注入
- 二分法高效定位问题

Reported-by credits in 5.4

Recipients		Creditors	
Hulk Robot	263 25.4%	Yue Haibing	236 22.8%
Syzbot	85 8.2%	Eric Dumazet	34 3.3%
kbuild test robot	71 6.8%	Gao Xiang	30 2.9%
Dan Carpenter	28 2.7%	zhengbin	25 2.4%
Stephen Rothwell	26 2.5%	Will Deacon	14 1.4%
Christoph Hellwig	25 2.4%	Chris Wilson	14 1.4%
Qian Cai	10 1.0%	Gustavo A. R. Silva	14 1.4%
Geert Uytterhoeven	9 0.9%	Jens Axboe	14 1.4%
Andrew Lunn	8 0.8%	Nathan Chancellor	13 1.3%
Arnd Bergmann	6 0.6%	Stephen Boyd	10 1.0%

完整的CVE安全漏洞管理

- 社区全量CVE跟踪
- CVE全生命周期管理和测试

终端领域创新 – 围绕1 + 8 + N策略为社区打造新终端存储

2015

终端存储1.0
F2FS

- 开源社区引进
- 强化稳定性
- 智能化空间回收
- 智能化Discard
- 增强修复能力

2018

终端存储2.0
EROFS

- 反哺开源社区
- 面向终端场景的磁盘格式
(极致精简、变长压缩...)
- 完善安全特性
- Inline数据
- 极致IO流程

2019

终端存储3.0
DFS

- 多设备数据融合
- 极致性能 (4x Samba)
- 面向不稳定网络环境的网络管理

终端领域开放创新案例 – 用户体验提升+ 开源 + 学术论文

More Speed, More Space

Previously, the F2FS file system was tailored for flash memory to improve the read and write performance after long-term use. Now, with **EROFS**, the random read speed will be 20% faster¹ and the system space will provide extra personal storage of up to 1000 pictures or 500 songs.² Plus, the read-only memory design isolates outside interference for added security in your system files.

Random read performance



Available space



Linux Kernel 4.19 Released with Initial Wi-Fi 6 Support, New EROFS File System

It's coming soon to a GNU/Linux distribution near you



EROFS: A Compression-friendly Readonly File System for Resource-scarce Devices

Xiang Gao, *Huawei Technologies Co., Ltd.*; Mingkai Dong, *Shanghai Jiao Tong University*;
Xie Miao, Wei Du, and Chao Yu, *Huawei Technologies Co., Ltd.*;
Haibo Chen, *Shanghai Jiao Tong University / Huawei Technologies Co., Ltd.*

<https://www.usenix.org/conference/atc19/presentation/gao>

This paper is included in the Proceedings of the
2019 USENIX Annual Technical Conference.
July 10–12, 2019 • Renton, WA, USA
ISBN 978-1-939133-03-8

Open access to the Proceedings of the
2019 USENIX Annual Technical Conference
is sponsored by USENIX.



- 1 / 华为在Linux内核的贡献历程
- 2 / 业务创新驱动开源
- 3 / 聚焦的方向和思考
- 4 / 小结

小结

- 华为积极拥抱Linux开源社区，是Linux社区的受益者，也逐步成为主要贡献者之一
- 通过围绕应用场景、硬件演进与技术创新等，积极在Linux社区创新
- 期待和CLK各位同事一起努力，共同推进Linux社区的演进



THANKS