



# 代际交错背景下移动蜂窝系统的近因现象与自调控设计: 速度、能耗与可靠性

刘云浩\*, 李振华, 李洋, 林灏, 宫良一, 郑雅文

清华大学信息学院, 北京 100084

\* 通信作者. E-mail: yunhao@tsinghua.edu.cn

收稿日期: 2022-02-09; 修回日期: 2022-05-02; 接受日期: 2022-06-17; 网络出版日期: 2022-12-07

国家科技部重点研发专项“工业互联网网络模态关键技术研究及验证”(批准号: 2021YFB2900100) 资助项目

**摘要** 移动蜂窝系统正经历着从 2G/3G/4G 到 5G 的交错式代际进化过程, 多样化的信号与协议长期共存. 如何在庞大、异构、高动态的蜂窝环境中保证服务的高速 (传输快)、高能效 (耗电少) 和高可靠性 (不断网) 成为至关重要但又非常困难的问题, 该问题同时涉及手机制造商、网络服务提供商、操作系统和应用软件开发商. 我们研究和综述了近年来针对该问题的代表性工作, 发现蜂窝系统表现层面的多个反常性能问题, 实际上归因于核心机制层面隐蔽而深刻的“近因现象”: 受社会舆论和商业宣传影响, 在选择使用移动网络时, 存在非理性的代际偏颇和不自觉的速度至上倾向, 进而引发了偏向 5G 的“亏损迁移”、网络容量的“饱和饥饿”以及基站连接的“乒乓跳跃”等连锁效应. 为修复上述缺陷, 我们提出基于 4G/5G 双连接以及非齐次时变马尔可夫过程的自调控技术设计, 将 5G 设备的故障发生率降低 40% 并将所有设备的平均故障持续时间缩短 36%, 同时探讨我们基于大规模工业合作实践的洞察和创新.

**关键词** 代际交错, 5G, 近因现象, 亏损迁移, 自调控技术

## 1 引言

移动通信技术在最近三十年的发展历程中, 依靠丰富的模式创新与快速的技术迭代, 在传输数据率<sup>[1]</sup>、覆盖范围<sup>[2,3]</sup>、功耗开销<sup>[4,5]</sup>三个核心维度不断拓宽与突破, 服务数百亿移动终端设备多样化的连接通信需求, 在很大程度上重塑了人类交流沟通的形态和方式. 其中, 作为当前移动通信覆盖范围最广、影响人群最多的普适技术, 蜂窝网络正经历从 2G/3G/4G 到 5G/6G 的关键进化; 如图 1 所示, 伴随这一进化过程, 作为基础传输架构的蜂窝网络所支撑的应用类型, 已经涵盖了前沿的增强/虚

**引用格式:** 刘云浩, 李振华, 李洋, 等. 代际交错背景下移动蜂窝系统的近因现象与自调控设计: 速度、能耗与可靠性. 中国科学: 信息科学, 2022, 52: 2290-2305, doi: 10.1360/SSI-2022-0062

Liu Y H, Li Z H, Li Y, et al. Recency effect and self-regulating design of mobile cellular systems in the context of interlaced generations: network bandwidth, power efficiency, and connection reliability (in Chinese). Sci Sin Inform, 2022, 52: 2290-2305, doi: 10.1360/SSI-2022-0062

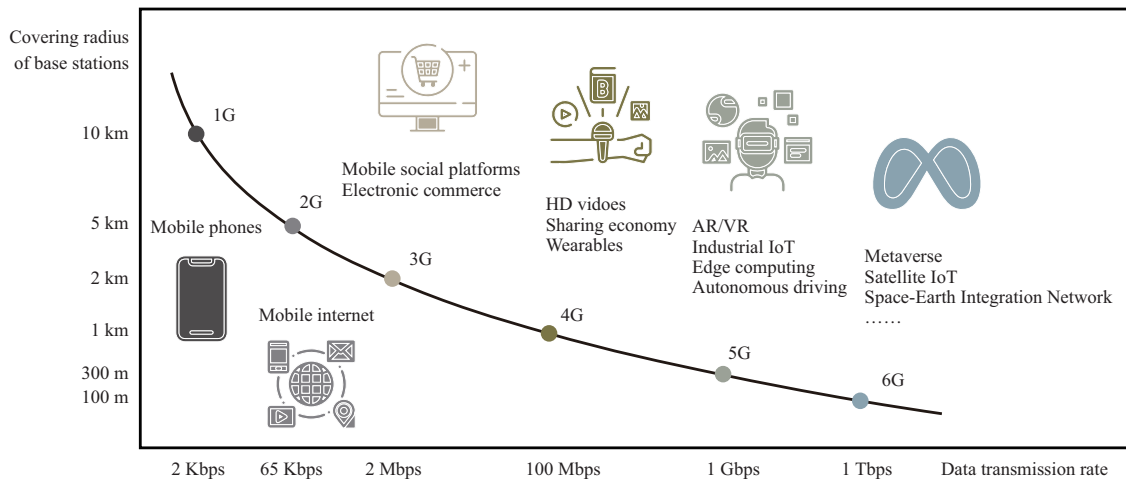


图 1 (网络版彩图) 移动蜂窝网络的代际进化及其支撑的应用类型

Figure 1 (Color online) Evolution of cellular networks and their supporting applications

拟现实 [6]、工业物联网 [7,8]、边缘计算 [9] 和自动驾驶 [10], 在不久的将来有望包括更具挑战性的元宇宙 [11]、卫星物联网 [12,13] 和天地一体化 [14] 等。

以 5G 为代表的高速蜂窝网络体现着当前移动通信技术的极端性能追求 [15,16]。人们对 5G 的一般理解是其具有高带宽的特点, 即数据下载速度达到 1 Gbps 以上, 甚至达到 10 Gbps; 相应的数据上传速度达到 100 Mbps 以上, 甚至 1 Gbps, 比 4G 快将近 100 倍。但实际上, 5G 的追求远不止于高带宽, 它还要求低时延 (毫秒级, 比 4G 小数十倍) 和高服务密度 (每平方公里百万部手机, 比 4G 大上百倍)<sup>1)</sup>。另一方面, 由于使用高频段无线信号、电波绕射能力弱, 单台 5G 基站的信号覆盖半径只有 100~300 m, 比 4G 小 10 倍左右, 手机移动时将需要更频繁地切换基站。

值得注意的是: 5G 的出现并不会立刻导致传统 2G/3G/4G 基站及对应手机通信功能的消失 (至少 5~10 年内不会), 多样化的信号与协议将长期共存, 形成“代际交错”的基本产业背景和生态系统<sup>2)</sup>。因此, 如何在庞大、异构、高动态的蜂窝环境中保证服务的高速 (传输快)、高效 (耗电少) 和高可靠性 (不断网) 成为至关重要但又非常困难的问题 [17]。该问题同时涉及手机制造商、网络服务提供商、操作系统和应用软件开发商, 一直是移动通信学术界和工业界热切关注但捉摸不定的焦点。

本文调研了近年来针对上述焦点问题的国内外代表性研究工作。其中既有小规模的前瞻型研究, 比如加利福尼亚大学伯克利分校 (University of California, Berkeley) 的学者对如何打破当前蜂窝市场的巨头垄断, 建立民主化的自由竞争秩序的讨论 [18], 以及惠普实验室的研究人员对极端动态环境下蜂窝基站交接方式的探究 [19]; 也有中规模的测量型研究, 比如北京邮电大学的学者在校园范围跨层测量 5G 网络的速度、能耗和覆盖范围 [20], 以及明尼苏达大学双城分校 (University of Minnesota, Twin Cities) 的学者在美国境内多角度测量 7 部 5G 手机的网速、用电和用户体验 [21]; 还有大规模的系统型研究, 比如普渡大学 (Purdue University) 的研究人员在世界范围内分析 30 家移动运营商的蜂窝基站交接情况 [22], 以及本文作者在全国范围内诊断和增强 7000 万部手机 (涉及三家移动运营商) 的蜂窝连接可靠性 [23]。

1) Clare Duffy. 2020. The big differences between 4G and 5G. <https://www.cnn.com/2020/01/17/tech/5g-technical-explainer/index.html>.

2) NepaliTelecom. 2020. Why Do Three Mobile Technologies 2G, 3G, 4G Co-exist? <https://www.nepalitelecom.com/2019/03/why-2g-3g-4g-mobile-technologies-coexist.html>.

深度分析上述研究总结提出的表现层面的各种反常性能问题, 我们发现, 这些问题实际上归因于蜂窝系统核心机制层面隐蔽而深刻的“近因现象”<sup>3)</sup>: 受当前社会舆论和商业宣传影响, 用户在选择使用移动网络时, 存在非理性的代际偏颇和不自觉的速度至上倾向, 这是损害蜂窝系统性能的主要根源. 比如北京邮电大学的研究者发现, 5G 在户外无遮挡的环境中速度很高, 但只要进入室内, 信号强度就下降一半以上; 并且对于普通的上网场景 (例如网页浏览), 5G 不仅没有什么明显优势, 相比 4G 反而多消耗 2~3 倍的电量, 体现出偏向 5G 的“亏损迁移”效应. 此外, 即使对于高速的上网场景 (比如增强现实), 5G 对其物理层信道带宽的利用率也仅有 32%, 体现出网络容量的“饱和饥饿”效应<sup>[21]</sup>. 这些效应在明尼苏达大学双城分校研究者的测量结果中都得到了直接确认或间接佐证, 说明“近因现象”具有普适性.

相对于速度和能效, 蜂窝系统的可靠性研究难度很大, 因为连接故障事件在实验室环境中极难复现<sup>[24]</sup>. 为了全面深刻地理解并切实有效地解决可靠性问题, 我们和小米手机团队密切合作, 大规模、长时间地研究影响蜂窝系统可靠性的软硬件及人类行为因素. 基于定制化操作系统, 连续八个月跟踪测量了 34 个型号、约 7000 万部小米手机的连接故障 (也称“异常断网”) 事件. 在二百多天时间里, 平均每部手机的异常断网事件多达 33 次, 每次断网持续时间长达 3.1 min. 通过阅读源代码, 我们发现这些问题很大程度上归咎于安卓操作系统蜂窝数据连接管理模块中的设计缺陷, 特别是偏向 5G 的“激进迁移”和发现连接故障后固定周期的“迟滞恢复”. 这样的现状还与移动运营商的基站部署策略紧密相关: 过度稀疏的基站布局固然容易引起手机断网, 但过度密集的基站布局也会导致数据连接的“乒乓跳跃”效应<sup>[25]</sup>, 使附近设备与不同基站频繁进行连接的建立与切断, 致使用户体验受损.

为修复上述技术缺陷, 研究者提出了一系列优化算法或数据驱动自调控创新设计. 比如惠普实验室的研究者提出基于手机运动模式的基站交接机制, 来替换经典的基于无线信号强度的基站迁移机制, 使得即使在极端的动态环境下也有望维持稳定的高性能蜂窝服务. 加州大学伯克利分校的研究者想法更为奇特, 设计出被称为“蜂窝砖头” (CellBricks) 的全新架构, 将蜂窝系统的移动支持和用户管理功能从网络端迁移到手机端. 此外, 北京邮电大学和明尼苏达大学的研究者都提出了一些基于统计数据的启发式方法, 能够帮助手机智能且无损地在 4G 和 5G 接入之间切换.

已有的这些创新设计的实际运行效果尚无把握, 它们的实验评估要么规模小、要么依赖大数据模拟. 得益于小米手机团队的支持, 我们设计实现并规模部署了两项新的设计: (1) 引入稳定性兼容的 4G/5G 双连接机制, 实现蜂窝连接的高可靠平滑切换; (2) 提出基于“非齐次时变马尔科夫过程”的最优化方法, 将被动等待的故障恢复策略革新为自适应的主动触发策略. 在 2800 万部小米手机上应用两个月后的实际效果表明: 5G 手机异常断网次数减少 40%, 所有手机异常断网持续时间缩短 36%. 一些初始结果发表的论文<sup>[23]</sup> 获得 ACM SIGCOMM 2021 唯一最佳学生论文奖, 相关代码已开源<sup>[26]</sup>.

正文部分组织结构如下: 首先, 第 2 节综述移动蜂窝系统领域近年来的代表性研究工作, 特别是我们从其中总结归纳出的“近因现象”. 其次, 第 3 节介绍我们为研究蜂窝系统可靠性问题而开展的大规模长时间测量工作及其主要科学发现. 之后, 第 4 节讲解我们为增强蜂窝系统可靠性提出的两项自调控设计及其实际应用效果. 最后, 第 5 节讨论移动蜂窝系统的未来发展方向.

## 2 代表性研究工作回顾与总结

代际交错的基本产业背景下, 移动蜂窝系统的异构性、复杂性与动态性剧增<sup>[27]</sup>, 如何为用户提供高质量的服务, 特别是如何保证服务的高速度、高能效和高可靠, 面临更多的困难和挑战, 相应地需要

3) ScienceDirect. 2022. Recency Effect. <https://www.sciencedirect.com/topics/psychology/recency-effect>.

更智能、更适配的技术设计甚至商业策略. 本节我们回顾近三年国内外代表性相关研究工作, 包括小规模的前瞻型研究、中规模的测量型研究和大规模的系统型研究, 以及我们对这些研究的归纳总结. 之所以选择“近三年”, 是因为2018年6月3GPP组织<sup>[28]</sup>发布了第一个5G标准 (Release-15<sup>[29]</sup>)、支持5G独立组网, 至今刚过三年.

## 2.1 小规模前瞻型研究

虽然小规模研究的结果往往不具备代表性, 但非常适合面向未来和极端场景的前瞻性探索<sup>[18, 19, 30~32]</sup>. 移动蜂窝产业和诸多产业一样, 如果存在多个组织、多套方案良性竞争, 则能让消费者最大程度受益, 同时亦鼓励创新. 遗憾的是, 这个产业的进入门槛很高、生存门槛更高, 对体量小的运营商或探索期的创新方案很不友好, 最终形成行业巨头垄断、技术方案固化的不利局面<sup>4)</sup>.

针对上述问题, 加州大学伯克利分校的学者从体系结构角度开展理论研究, 讨论如何打破当前蜂窝市场的巨头垄断, 建立民主化的自由竞争秩序, 让创新但尚不成熟的方案拥有更公平的机会. 具体地, 他们设计出称为“蜂窝砖头”(CellBricks)的全新体系结构<sup>[18]</sup>, 在蜂窝系统的技术栈中, 将最重要的移动支持和用户管理两大功能从网络端迁移到手机端, 使得用户能够自主访问身边的任何基站、不被特定运营商锁定. 该体系结构还能使蜂窝基础设施更简单、频谱资源使用更高效, 当然也会带来一系列政策和商业问题.

移动蜂窝系统设计之所以复杂, 最大的挑战是终端设备的“移动”性<sup>[33]</sup>, 其极端场景是时速高达300公里的高铁网络. 为此, 惠普实验室的研究人员专门探究5G (甚至超越5G)时代的极端移动性管理<sup>[19]</sup>. 通过对现存蜂窝基站交接机制的深度分析, 他们认为, 基于无线信号强度的经典思维虽然适合于静态和低速场景, 但对于高速场景下基站交接行为的触发、决策和执行无法令人满意, 因为此时的多普勒频移效应<sup>[34, 35]</sup>已经严重到无法忽视的程度. 另一方面, 他们发现高速场景下的手机运动实际上更为规律、更可预测, 从而提出基于手机运动模式的基站交接机制, 在延迟多普勒域<sup>[36]</sup>构造信号覆盖, 使用新出现的正交时频空间调制技术<sup>[37]</sup>提取蜂窝终端运动以及多径信息. 小规模仿真和大规模高铁数据模拟都表明, 这一新机制有望将高速场景下的基站交接可靠性提升到原先静态和低速场景的水平.

## 2.2 中规模测量型研究

为克服小规模研究的缺点, 同时受限于高校科研人员在平台系统方面的不足, 中规模的测量型研究被高校师生广泛开展. 虽然规模受限, 但在蜂窝系统测量方法、部署方案、数据处理方面都能够达到较高的自由度和较细的粒度, 取得具备说服力的结果<sup>[20, 21, 38~41]</sup>.

基站是蜂窝系统的核心, 但以基站为中心的测量研究十分稀缺, 因为基站数据极难获取, (被运营商)高度保密. 为填补这一空缺, 北京邮电大学的研究者在不到一平方公里的校园范围内, 通过专业设备和专门软件长时间测量6台5G基站的覆盖范围、速度和能耗, 研究层面包括信号层、网络层、传输层和应用层<sup>[20]</sup>. 他们发现5G在户外的无遮挡环境中速度很高, 但只要进入室内, 信号强度就下降一半以上. 并且对于普通上网场景 (例如浏览网页), 5G不仅没有什么明显优势, 反而相比于4G将多消耗2~3倍的电量. 此外, 即使对于高速上网场景 (比如增强现实), 5G对其物理层信道带宽的利用率也仅有32%. 他们提出使用启发式方法将特定类型数据传输从5G迁移到4G, 预期能够省电约四分之一.

4) Emily Stewart. 2020. America's Monopoly Problem, Explained by Your Internet Bill. <https://www.vox.com/the-goods/2020/2/18/21126347/antitrust-monopolies-internet-telecommunications-cheerleading>.

与上述工作不同但形成互补的是, 明尼苏达大学双城分校的学者在美国境内租用 157 台地理分散的虚拟机服务器, 从多角度测量 7 部 5G 手机的网速、用电和用户体验. 其测量结果直接确认或间接映证了北京邮电大学学者的发现<sup>[21]</sup>. 所增补的测量发现: 关于独立工作的 5G 基站, 由于控制平面不再依赖 4G 协议, 其性能有明显提升. 此外, 对于自适应视频流的传输, 由于 5G 信号有限的传播能力, 传输阻塞故障经常出现. 对于用户最常用的网页访问, 虽然 5G 总体上能缩短加载时间, 但不同类型网页之间的差异巨大, 不过这种差异用决策树模型<sup>[42]</sup>辅助, 即可在 4G 或 5G 接入上做出明智选择.

### 2.3 大规模系统型研究

尽管中小规模研究有其独特的优点, 但都无法替代大规模系统型研究的真实性与说服力. 开展此类研究的难度很大, 一般需要移动运营商、手机制造商或操作系统开发商的支持. 如果缺乏底层硬件或基础软件提供者的支持, 通常只适合大规模研究移动蜂窝系统的某一个层面, 同时需要以众包的方式招募大量志愿者, 投入极大的时间与精力<sup>[22, 23, 43~45]</sup>.

前面我们介绍过惠普实验室的研究人员对蜂窝基站交接机制的深度分析, 但其仅对国内特定路线的铁路数据进行了仿真实验, 规模较小. 为克服这一缺点, 普渡大学的研究人员开发了专门的软件 MMLab, 招募了 35 名来自全球各地的专业志愿者, 在世界范围内广泛测量分析 30 家移动运营商的蜂窝基站交接策略<sup>[22]</sup>. 他们发现不同移动运营商所制定的交接策略差异很大并且都很复杂, 比如 4G 基站交接涉及多达 66 个配置参数, 而 2G/3G 基站交接则涉及 91 个配置参数; 即使是同一家运营商的同类型基站, 其配置参数的具体取值也会随信号频段以及基站位置而改变. 运营商通常都指望依靠多部署基站、增大信号强度来提升交接性能, 但实际上, 能够恰当地掌控交接时机对提升性能更为关键. 很可惜, 研究揭示的现状是交接参数经常配置不合理, 交接时机似乎也不(接近)最优.

在小米手机团队的支持下, 我们在全中国范围内大规模、长时间测量 7000 万部安卓手机(涉及三家移动运营商)的蜂窝连接可靠性, 揭示出多个学术界和工业界未知的技术缺陷, 提出并实践应用了两项自调控创新设计, 这些内容将在第 3 和 4 节分别加以详述.

### 2.4 小结

以上调研的各项工作, 虽然规模不同、各有侧重, 但都呈现出移动蜂窝系统表观层面的各种反常性能问题. 将这些表观问题聚集到一起梳理凝练、深度分析, 我们发现它们都能直接或间接归因于蜂窝系统核心机制层面隐藏而深刻的“近因现象”.

“近因现象”(又称“近因效应”)起源于用户心理学, 被定义为人对时序上最近的信息总能留下更深刻印象的现象, 该现象曾被巴黎大学(University of Paris)的研究者通过实验所证实: 当过时信息和最新信息同时在记忆中被激活时, 只有最新信息会对人的内容理解产生影响<sup>[46]</sup>. “近因现象”在多个面向用户的研究领域都受到广泛关注. 在体验质量(quality of experience, QoE)领域, 近因效应是用于量化用户体验与行为成因的一个重要效应. 德国维尔茨堡大学(University of Würzburg)的研究者在观察影响体验质量建模和管理的新要素时提出, 由人类短期记忆和记忆刺激能力引起的近因效应, 在基于互联网的各种传播质量随时间波动的多媒体信息(包括音频、视频)中都有体现<sup>[47]</sup>; 清华大学人工智能研究所的学者在建立针对搜索会话的更优评价指标时, 结合近因效应描述了用户在会话中的认知过程<sup>[48]</sup>; 华为研究团队在制定 5G 云 VR 服务体验质量标准时, 所用的测量机制之一就是“近因效应”<sup>[49]</sup>. 在其他工程领域的研究中, “近因效应”也常被用于解释某种后果主要由新近的信息所引发的现象: SAS 软件研究所(SAS Institute Inc.)的研究员以“近因效应”来解释电力需求受到最近几个小时温度影响的规律<sup>[50]</sup>; 中国人民解放军陆军工程大学(Army Engineering University of PLA)的学

者用“近因效应”阐释在经典塑性力学的加载路径问题中,后加载的应变对最终应力状态影响更大的现象<sup>[51]</sup>.

在本文中,“近因现象”则指用户受当前社会舆论和商业宣传影响(特别是对 5G 的表述与过度期待),在选择使用移动网络时存在非理性的代际偏颇和不自觉的速度至上倾向.这种“近因现象”成为损害蜂窝系统性能的主要根源.所调研的代表性研究工作基本来自近三年网络通信领域的主流学术会议(如 SIGCOMM, MobiCom 和 MobiSys),其中大概一半来自中国学者,另一半来自美国学者,说明“近因现象”是普适的,同时也反映出我国学术界在移动蜂窝领域科研能力的迅速提升.

### 3 大规模蜂窝系统的可靠性测量

蜂窝网络不断提升的服务速度、时延和覆盖范围指标都依赖于数据连通的前提,但针对蜂窝数据连接可靠性的研究却并未引起科研人员足够多的重视.依据移动设备与基站建立连接过程中的不同阶段,安卓操作系统将蜂窝系统的可靠性问题分为以下三类:“数据业务建立失败”<sup>5)</sup>,指移动设备从附近基站接收到信号,但是无法成功建立数据连接;“无数据服务”<sup>6)</sup>,指移动设备成功建立数据连接,但是无法收到蜂窝数据;“数据阻塞”<sup>7)</sup>,指移动设备已经成功建立数据连接并收到蜂窝数据,但在使用过程中超过 1 min 未接收到 TCP 数据包.为了更加深入地探讨蜂窝系统的可靠性,我们与安卓手机制造商小米公司合作,在设备终端侧对该问题进行了长时间、大规模的研究.

#### 3.1 测量方法

在采集数据前,我们发现安卓操作系统的蜂窝连接管理模块<sup>8)</sup>在实现上存在明显的设计缺陷:安卓操作系统没有为底层蜂窝连接组件提供足够的跟踪和日志记录,使得管理模块无法精确捕获故障事件,并有效地诊断其根本原因.这一缺陷从根本上损害了蜂窝连接的可靠性,同时也对准确测量可靠性故障事件造成了困难.由于安卓蜂窝模块测量设施<sup>8)</sup>的局限性,我们无法仅通过开发一个用户空间应用程序来采集蜂窝连接数据,而必须尝试在原生安卓系统之上构建定制化的蜂窝连接持续监控设施.如图 2 所示,为了捕获真正的故障事件并收集诊断信息,我们将监控设施注册为安卓蜂窝连接管理服务的事件监听器,通过检测相关的系统服务,来记录详细的基站与网络状态信息,以及部分故障的协议错误代码<sup>9)</sup>.

在开发过程中我们还发现,安卓系统在检测数据阻塞故障事件的过程中存在设计缺陷.如前文所述,当移动设备在 1 min 内发出超过 10 个出站 TCP 包,同时没有接收到任何入站 TCP 包时,安卓系统将该事件定义为“数据阻塞”.由其检测方式可知,该类型故障事件持续时间的测量误差最长可达 1 min.同时,在测量过程中缺乏关键的网络堆栈信息,对故障事件的误报概率较大.为修复这一缺陷,我们设计了轻量级的主动探测组件.当安卓系统第一次检测到数据阻塞事件时(这一次的检测使用安卓原本的方法),该组件将向本地 IP 地址发送 ICMP 信息,以检测本地网络栈的连通性;同时,该组件也

5) Android.org. 2021. Data Setup Error in Android. <https://android.googlesource.com/platform/frameworks/opt/telephony/+ /refs/heads/master/src/java/com/android/internal/telephony/dataconnection/DcTracker.java>.

6) Android.org. 2021. State Out of Service in Android. <https://developer.android.com/reference/android/telephony/ServiceState>.

7) Android.org. 2021. Data Stall Report in Android. <https://developer.android.com/reference/android/net/ConnectivityDiagnosticsManager.DataStallReport>.

8) Android.org. 2021. Data Connection Management in Android. <https://android.googlesource.com/platform/frameworks/opt /telephony/+ /master/src/java/com/android/internal/telephony/dataconnection/DataConnection.java>.

9) Android.org. 2021. Data Fail Cause in Android. <https://android.googlesource.com/platform/frameworks/base/+ /master/telephony/java/android/telephony/DataFailCause.java>.

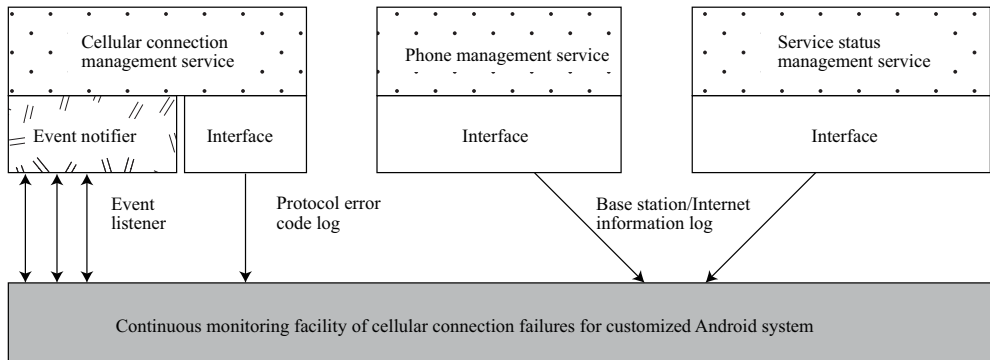


图 2 在安卓系统中捕获蜂窝连接故障的持续监控设施

Figure 2 Continuous monitoring infrastructure for capturing cellular failures in Android

会向设备分配的 DNS 服务器发送 ICMP 信息和 DNS 查询, 来确定连接中断问题是否是 DNS 域名解析服务不可用而导致的. 如果探测结果表明数据阻塞没有被修复, 我们将启动一个新的探测过程; 否则我们会把之前探测过程记录的持续时间相加来估计该数据阻塞事件的持续时间. 此外若事件持续时间超过 1200 s (仅不到 10% 的事件), 该组件则会从下一个探测过程开始每次探测将超时时间翻倍来减少测量开销. 相较于原生安卓系统的测量方式, 该组件使用的发送 ICMP 信息及 DNS 查询操作的超时时间均在 5 s 以内<sup>[52,53]</sup>, 大幅降低了对数据阻塞事件持续时间的测量误差, 并可以有效地过滤误报信息.

基于上述两个方面的优化, 我们将蜂窝连接持续监控设施和数据阻塞主动探测组件集成到原生安卓系统, 与小米公司合作开发了定制化的安卓系统. 在定制化过程中, 我们专注于修改安卓系统框架层程序, 以提供轻量级的、保护隐私的细粒度数据跟踪和分析. 我们还对定制化后的系统开销进行了测试, 结果表明即使是目前的低配置安卓手机, 在检测到的故障持续时间内, 我们的修改平均也只产生了额外的 <2% 的 CPU 利用率, <40 KB 的内存使用量, 以及 <100 KB 的存储空间; 每月的网络使用量 <100 KB. 在没有手机故障的日常使用中, 监控基础设施在客户端处于休眠状态, 因此不会产生额外的开销. 另一方面, 定制化测量所产生的网络开销即使累加在一起也相当低: 对于所有参与研究的用户设备, 每秒钟的总网络开销均低于 500 KB, 因此对性能的影响可以忽略不计. 此外为了不产生额外流量使用, 记录的数据只有在有 WiFi 连接时才会上传到后台服务器.

在测试过程中, 我们通过系统更新提示向 2.5 亿小米手机用户发送了合作研究邀请, 充分告知用户此次研究的目的、开销与细节. 如果用户接受邀请, 则可以在手机上安装定制化系统来参与研究. 最终, 共有 7000 万用户选择参与研究, 并与我们共享了 8 个月 (从 2020 年 1 月至 2020 年 8 月) 的数据, 数据集涉及三大移动设备运营商、34 款不同型号的安卓手机和 500 多万基站. 表 1 详细给出了数据集中每条记录的具体属性, 其中不包含任何安全或隐私敏感信息.

### 3.2 数据分析

观察整体的数据分布, 我们发现, 蜂窝连接故障事件的实际严重性, 比用户之前反馈的情况还要糟糕. 在测量过程中, 平均每部手机的异常断网事件数多达 33 次, 每次断网事件的持续时间长达 3.1 min, 具体分布如图 3 和 4 所示. 设想在紧急情况下, 用户无法在 3 min 内与外界取得联系, 这个用户体验很差, 甚至有可能导致严重的后果. 我们从基站部署、硬件配置和操作系统等角度, 对造成蜂窝故障事件的根本原因进行了如下分析.

表 1 测量中采集到的蜂窝系统可靠性数据集的具体属性

Table 1 Attributes of the cellular reliability dataset collected in the measurement

Attribute	Description
UID	Unique ID generated to identify a user (cannot be related to the user's true identity)
TIME	UNIX timestamp
FAILURE.TYPE	Data_Setup_Error, Out_of_Service, and Data_Stall
FAILURE.DURATION	Duration of cellular failures (mainly contributed by Data_Stall)
RAT	Current radio access technology (2G/3G/4G/5G)
RSSI	Signal strength in dBm
CELL	Base Station ID (MCC+MNC+LAC+CID)
OS	Android version (9/10)
MODEL	Device model
CAUSE	Error code of Data_Setup_Error defined by 3GPP [28]
APN	Current access point names

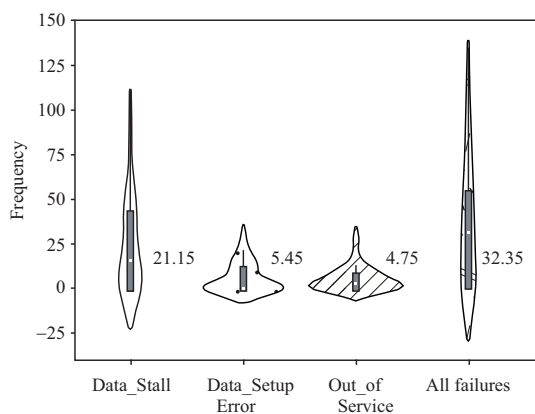


图 3 平均每台手机上各蜂窝故障事件的发生次数

Figure 3 Average number of cellular failures experienced per phone

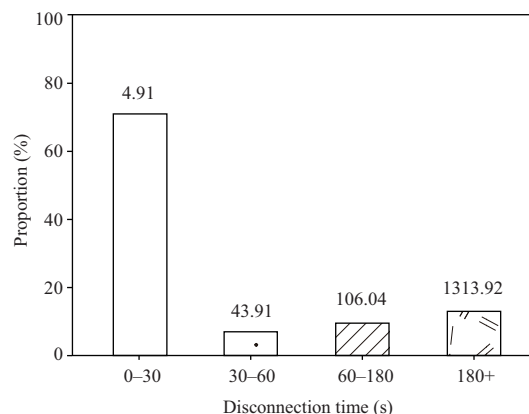


图 4 捕获到的蜂窝故障事件的持续时长分布

Figure 4 Distribution of the recorded cellular failure events' duration

### 3.2.1 基站部署

从基站发送信号的角度看, 人们的普遍认知是, 蜂窝可靠性与接收信号强度呈正相关. 然而, 我们的测量结果却与此不符. 从数据来看, 相较于信号显示 1~4 格的较弱信号情况, 蜂窝连接故障更容易发生在信号显示 5 格 (满格) 的高质量信号情况下. 通过仔细对比高信号强度下故障事件的各个数据维度, 我们发现, 大多数此类故障都来自于公共交通枢纽周围密集部署的基站. 虽然这种部署策略能够提供高信号强度, 但也大大增加了在基站间进行切换的控制信道开销 [54, 55]. 甚至在极端情况下, 即使周围基站的服务效果都比较好、没有本质区别, 一部手机也可能在很短的时间内反复切换连接的基站. 这种现象被形象地称为“乒乓跳跃”效应 [25]. 在这样的情况下, 不同运营商的相邻基站之间也会存在严重的邻频干扰 [56], 导致蜂窝连接故障发生的隐患大幅增加.

从基站侧观察不同代际服务的蜂窝连接可靠性, 我们发现 3G 基站的故障率 (10.2%) 明显低于 2G 基站 (23.4%) 和 4G 基站 (65.2%). 进一步分析测量数据, 我们发现, 这是由移动设备的连接优先级和



不同代际基站的特性差别导致的: 在 4G 服务可用时, 用户手机一般会与 4G 基站建立连接. 如果 4G 服务不可用, 由于 2G 基站的信号覆盖范围明显大于 3G 基站<sup>10)</sup>, 绝大多数的用户手机会与 2G 基站建立连接. 这一移动设备的连接策略会引发蜂窝网络容量使用情况的不平衡, 使得 2G 和 4G 基站繁忙, 而 3G 基站却相对空闲, 从而进一步导致与 3G 基站关联的蜂窝故障事件较少. 因为用户手机倾向于连接网络容量相对饱和的基站, 移动设备终端在竞争电信资源的过程中将更容易出现断网事件. 这种现象被称为蜂窝网络容量的“饱和饥饿”效应<sup>[21]</sup>.

### 3.2.2 硬件配置

除了基站连接层面的因素, 用户使用的不同移动设备也会对蜂窝网络的可靠性造成影响.

在硬件配置层面, 人们普遍的看法是, 更好的硬件应该为用户带来更稳定可靠的数据连通体验. 但是, 我们在测量过程中惊讶地发现, 随着硬件配置的提升, 蜂窝故障的发生概率和频率都趋于增加. 我们观察到, 至少经历过一次蜂窝连接故障的设备数量占其型号设备总数的比例平均为 23%, 依型号不同最低为 0.15%, 最高达 45%. 我们对频繁发生蜂窝故障事件的手机型号进行观察对比后发现: 问题严重的机型大多支持 5G. 故而我们推测, 这一现象很有可能是由 5G 模块的高工作负载<sup>11)</sup>以及当前相对不成熟的生产状态导致的.

### 3.2.3 操作系统

在用户手机的系统软件层面, 我们通过分析数据, 进一步发掘出了安卓操作系统蜂窝管理模块的多处隐式设计缺陷. 这些缺陷, 一方面使得蜂窝故障的发生概率增加, 另一方面使得蜂窝故障的持续时间增加, 均对蜂窝网络的可靠性造成了损害.

首先我们观察到, 安卓操作系统较新的版本 (例如安卓 10) 中追加了 5G 相关的新特性<sup>12)13)</sup>, 当安卓 10 系统在设备能接收到 5G 信号时, 会优先选择 5G 连接. 即使在 5G 信号远弱于其他无线接入技术的信号时, 安卓 10 仍然会优先选择 5G 连接. 由于该过程完全忽视了信号质量对数据连通可靠性的重要影响, 我们称之为偏向 5G 的“激进迁移”. 这一缺陷明显增加了蜂窝故障事件的发生率.

通过分析用户在发生蜂窝故障事件后的操作行为数据, 我们还发现, 不同版本的安卓系统针对蜂窝故障的自发恢复机制都存在设计缺陷. 具体来说, 对于故障持续时间最长 (占有事件总持续时间的 94%) 的数据阻塞问题, 安卓系统采用了一种三阶段渐进式的恢复机制. 在检测到数据阻塞事件后, 依次尝试轻量级 (清理并重新启动当前连接)、中等量级 (重新注册到网络) 和重量级 (重新启动蜂窝组件) 的恢复操作. 在采取每个操作之前会间隔一分钟, 以检测和判断数据阻塞问题是否仍然存在. 然而, 我们收集到的用户断连恢复行为数据显示, 对于大多数的数据阻塞问题, 如果移动设备无法在 30 s 内完成自动修复, 用户就会选择手动重启蜂窝数据连接, 从而直接中断原生安卓系统的三阶段恢复过程. 由于这一机制在实际应用中操作执行有效, 但触发间隔滞后低效, 我们称之为周期固定的“迟滞恢复”. 这一缺陷会使得操作系统不能及时地协助用户自动修复连接故障, 从而导致蜂窝故障的持续时间增加.

10) GSM. 2022. Global System for Mobile communications (GSM). <http://4g5gworld.com/wiki/global-system-mobile-communications-gsm>.

11) Brandon Lewis. 2020. 5G Workloads Demand More Horsepower. <https://www.insight.tech/content/5g-workloads-demand-more-horsepower>.

12) Android.org. 2021. Android 10 Highlights. <https://developer.android.com/about/versions/10/highlights>.

13) Android.org. Enhance Your Apps with 5G. 2021. <https://developer.android.com/training/connectivity/5g/enhance-with-5g>.

## 4 大规模蜂窝系统的可靠性增强

第3节讨论了安卓操作系统的蜂窝管理模块存在的多处重要的隐式设计缺陷. 针对这些缺陷, 我们设计了相应的修复机制来增强数据连接的可靠性, 以降低故障事件的发生率、缩短故障的持续时间, 并通过大规模的工业部署检验其实际效果. 我们向第3节提到的测试用户推送了添加修复补丁的新版定制化安卓系统. 2020年下旬, 该系统被40%的参与用户(约2800万)接受并安装更新.

### 4.1 降低故障发生率

针对安卓操作系统偏向5G的“激进迁移”现象, 我们着重处理了5G在0格信号下蜂窝故障高发的情况, 避免由较强4G信号向弱5G信号的切换. 经优化后的无线接入技术切换策略, 不再盲目追求5G潜在的高带宽, 可以降低因低信号质量和“乒乓跳跃”式切换导致的高故障发生率.

不同网络制式和信号强度下设备的蜂窝故障发生率数据表明, 4G在1~4格信号强度下的故障率远低于5G在0格信号下的故障率, 其中1~2格时低30%, 3~4格时低37%. 在这种情况下, 如果从4G向5G网络切换, 由于切换后的信号强度非常弱, 设备的数据速率和网络性能很难有正面提升, 数据连接的稳定性反而会下降. 因此, 我们在设计定制化安卓系统时, 规避了这类会大幅提高蜂窝故障风险的无线接入技术的自动切换. 同时, 我们还利用新颖的4G/5G双连接机制, 来优化安卓10中的无线接入技术选择策略<sup>[57]</sup>. 4G/5G双连接已成为从4G向5G代际过渡直至5G全面部署过程中的关键技术, 其理论基础与可行性也是蜂窝通信领域的研究热点. 意大利帕多瓦大学(Università degli Studi di Padova)的学者将上行控制信号系统与本地协调器相结合, 提出了4G/5G双连接协议, 并已实际应用于预防毫米波蜂窝系统的链路阻塞问题, 在链路故障时能够自动切换链路<sup>[58]</sup>; 韩国世宗大学(Sejong University)的学者综述了该技术在5G普及过程中的实现方案、挑战和未来方向, 同时评估了不同方案的性能<sup>[59]</sup>. 上述工作表明了4G/5G双连接技术在代际交错背景下确保蜂窝可靠性的优势与可行性, 因此我们将该技术优化并应用于移动设备的4G/5G网络切换: 在控制层面, 我们使设备同时维持与4G和5G基站的连接; 在数据层面, 则仅维持主连接的工作. 在这种策略下, 设备可以更稳定平滑地切换无线接入技术, 降低蜂窝连接故障发生的隐患.

如图5所示, 截至2020年12月底, 在不影响所有参与用户手机数据速率的前提下, 我们设计的稳定兼容RAT切换成功减少了40%的5G手机蜂窝连接故障事件. 图中数据表明, 针对安卓系统定义三类故障事件, 数据业务建立失败、无数据服务和数据阻塞故障的发生频率分别降低了25.72%, 42.4%和50.26%, 其中发生最普遍的数据阻塞事件受到的优化效果尤为显著. 我们认为, 安卓偏向5G的“激进迁移”设计会使移动设备经常在4G, 5G基站间频繁切换连接(“乒乓跳跃”效应), 增加了数据阻塞故障发生的可能性. 而我们对4G/5G高隐患切换场景的规避, 以及对平滑的双连接机制的利用, 能有效解决这一问题.

### 4.2 缩短故障持续时间

针对安卓操作系统固定周期的“迟滞恢复”现象, 考虑到系统的恢复操作有效而触发间隔滞后, 我们希望找到最优的恢复操作触发间隔, 来增加系统自动修复蜂窝连接故障的可能性, 以缩短蜂窝故障的持续时间.

为了优化数据阻塞故障的修复过程, 我们将该故障事件的整个恢复过程建模为一个马尔可夫过程<sup>[60]</sup>. 传统马尔可夫过程是状态转移概率仅与前后状态相关的稳态过程. 但在我们的问题场景下, 状态转移概率将根据在不同时间点下修复故障的用户比例来近似估计. 也就是说, 状态转移概率还与时

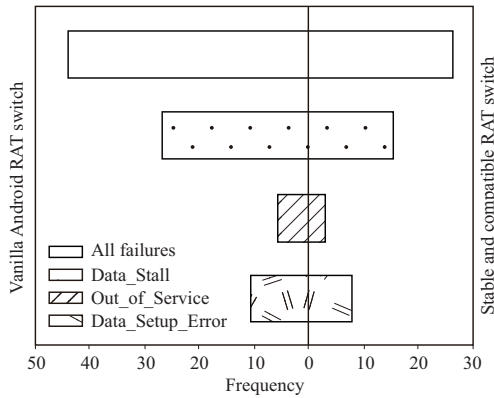


图 5 原生安卓 RAT 切换与稳定兼容 RAT 切换下的蜂窝故障事件发生频率对比

Figure 5 Frequency of cellular failures with the RAT transition policy of vanilla Android and the stability-compatible RAT transition policy

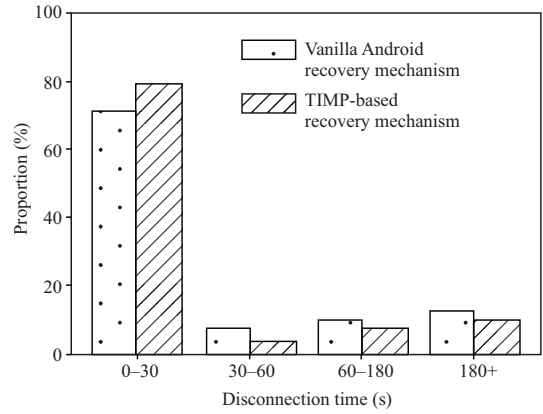


图 6 原生安卓恢复机制与基于 TIMP 的恢复机制的蜂窝故障时长分布对比

Figure 6 Duration of cellular failures with the Data\_Stall recovery mechanism in vanilla Android and the TIMP-based recovery mechanism

间变量相关. 因此, 我们无法直接采用传统的马尔可夫模型对恢复过程进行建模.

考虑到上述问题, 我们开发了基于非齐次时变马尔可夫过程 (time-inhomogenous Markov process, TIMP) 的灵活动态触发器<sup>[60]</sup>, 取代原生安卓系统中现有的固定时间间隔触发器, 从而修复 3.2.3 小节详述的三阶段渐进式蜂窝连接恢复机制中存在的缺陷. 相比于只能对静态过程进行建模的传统马尔可夫过程, 非齐次时变马尔可夫过程能以时间敏感的方式对复杂的状态转换过程进行建模. 此外, 意大利萨莱诺大学 (Università degli Studi di Salerno) 的学者曾通过模拟 Prendiville 过程, 从数学的角度证明了非齐次时变马尔可夫过程用于故障恢复的可靠性<sup>[61]</sup>. 如式 (1) 所示, 在非齐次时变马尔可夫过程中, 状态  $i$  与  $j$  之间在特定时间节点  $t_k$  的转移概率, 可以由该时间节点下的状态转移发生概率矩阵 ( $G_{i \rightarrow j}(t_k)$ ) 和状态转移接受概率矩阵 ( $A_{i \rightarrow j}(t_k)$ ) 共同决定. 基于此, 我们对整体的故障恢复过程进行建模, 使用退火算法<sup>[62]</sup> 来搜索全局最小值, 从而将三个机制触发间隔自适应地最优化为 21 s, 6 s 和 16 s, 相比原生安卓三个 1 min 的低效滞后的触发间隔有大幅度缩减.

$$\mathbb{P}_{(i \rightarrow j)}(t_k) = \begin{cases} G_{i \rightarrow j}(t_k)A_{i \rightarrow j}(t_k), & j \neq i, \\ 1 - \sum G_{i \rightarrow j}(t_k)A_{i \rightarrow j}(t_k), & j = i. \end{cases} \quad (1)$$

在实际部署中, 参与用户在研究期间的数据阻塞持续时间平均减少了 38%, 总故障持续时间减少了 36%. 从图 6 可以看出, 数据阻塞事件恢复时间超过 30 s 的情况有明显减少, 在 30 s 内就能快速恢复的设备比例明显提高: 断网时长在 30 s~1 min、1~3 min 和 3 min 以上的用户比例分别减少了 39.2%, 25.2% 和 19.2%, 在 30 s 以内快速恢复的用户比例提高了 10.5%. 这都得益于基于非齐次时变马尔可夫过程的数据阻塞恢复机制的优化. 尤为重要的是, 考虑到未来安卓系统可能出现的行为模式变更, TIMP 恢复机制能以灵活自适应的方式执行并优化蜂窝可靠性.

## 5 讨论与前瞻

移动蜂窝系统作为移动生态基础架构的一部分, 已然成为一项无处不在、影响深远的综合技术.

作为关键基础设施,移动蜂窝系统承载着工业互联、边缘感知和网络经济等重要新兴国民生产要素.因此,对移动蜂窝系统技术优势与缺陷的深刻理解、有效诊断和切实修复,是学术界与工业界均不可忽视的基础课题.本文由此出发,深入结合代际交错的复杂产业背景,在多个层次(终端设备、终端操作系统、蜂窝基站和移动运营商<sup>[63,64]</sup>)的代表性学术成果和工业产品中研究移动蜂窝系统的相关问题.

研究揭示出这一现代移动生态的重要支撑部分在网速、能耗、可靠性等方面存在普发且严重的未知问题,特别是发现受制于“近因现象”的思维定势,强大的社会舆论和商业宣传明显影响了代际交错背景下蜂窝系统基础软件的理性设计.在终端侧主要表现为不适应新式网络高吞吐、低覆盖特性的网络连接管理模块,而在网络侧主要表现为异构兼容性网络体系中存在的旧式网络缺陷.更重要的是,这些问题虽然发源于软件设计缺陷,且长时间地对用户造成负面影响,却往往需要真实复杂的环境以及人为因素予以触发.这使得它们潜藏在基础移动通信网络架构中多年而未能被真正发掘,也导致其迟迟未能得到有效修复.值得注意的是,上述问题并非偶然存在于基于安卓系统的移动端,我们对基于iOS系统的移动端也进行了调研,发现在蜂窝网速方面其与基于安卓系统的移动端差别不大<sup>[65]</sup>,在5G能耗方面其表现甚至更差<sup>[14]</sup>;而在蜂窝可靠性方面,尽管iOS系统在14.1版本中更新的“智能数据模式”功能相比安卓5G的“激进迁移”在RAT切换方面看上去更稳定可靠<sup>[15]</sup>,但据用户反馈,新版的iOS 15的蜂窝网络连接稳定性很差,存在数据阻塞问题,甚至会出现完全无法访问蜂窝数据的情况<sup>[16]</sup>.针对上述现象,我们使用载有最新的iOS 15系统的移动设备进行了小规模测试实验,实验过程中我们发现测试设备在遭遇数据阻塞后的30 s内并没有进行自动重启蜂窝数据等尝试.这表明iOS系统也并没有有效的数据阻塞恢复机制,一定程度上证实了我们的测量揭示的问题也部分存在于基于iOS系统的移动端.

针对上述移动蜂窝系统的重要隐藏缺陷,本文从算法智能和数据理性出发,在代际交错的时代背景下研究适应新式网络架构的自调控设计,并结合工业实践将相关设计予以应用部署,从而实现了对数千万终端设备以及通信基站可靠性的全方面提升,修复了关键的系统缺陷和漏洞,大幅度降低了手机的断网时长和频率,相关成果在学术界以及工业界均得到正面反馈.

移动通信技术与很多传统技术在多个维度上有着显著的区别,包括网络带宽、时延<sup>[66]</sup>、服务密度、覆盖半径<sup>[67]</sup>、能耗比<sup>[68,69]</sup>以及移动性<sup>[70]</sup>等.随着移动通信技术的快速迭代,不同代际的技术在这些维度层面上呈现出完全不一致的涨落态势,诸如5G网络等新兴技术在带来更高的带宽、更低的时延以及更高的服务密度的同时,也在其他如覆盖半径、能耗比以及移动性等关键维度上做出了不小的妥协<sup>[17]</sup>.然而,这种涨落趋势却与从传统通信技术中发展而来、高异构兼容的基础软件设计不可避免地产生了严重冲突,加之相关问题的隐式形态进一步阻碍了小规模快速缺陷诊断与修复,在综合因素的影响下,这样的冲突导致新网络时代的基础体验在可靠性层面上不仅未得到明确改进,甚至产生了一定程度的退化.当前,这一冲突随着通信技术的迭代加速开始变得愈加令人无法忽视,尤其阻碍了架构在这些关键基础软件上(如自动驾驶<sup>[10]</sup>、智慧城市<sup>[71]</sup>以及智能制造<sup>[72]</sup>等新应用场景)的成熟落地.这些问题不仅是对软件设计的新挑战,更是对未来移动通信技术标准的新要求.在当前

14) Hartley Charlton. 5G Drains iPhone 12 Battery 20% Faster Than 4G in Benchmark. <https://www.macrumors.com/2020/10/21/iphone-12-5g-battery-drain/>.

15) Technology For You. 2020. iOS 14.1 makes 5G only a little less confusing than Android 11. <https://www.technologyforyou.org/ios-14-1-makes-5g-only-a-little-less-confusing-than-android-11/>.

16) GEEKSMODO STAFF. 2022. iOS 15: Cellular Data Not Working On iPhone? Fix! <https://geeksmodo.com/how-fix-cellular-not-working-on-iphone/>.

17) John Fruehe. Top 5G limitations and challenges for businesses. 2020. <https://www.techtarget.com/searchnetworking/tip/Top-5G-limitations-for-businesses-include-lack-of-range-devices>.

加速部署 5G、前瞻布局 6G<sup>[73]</sup> 的形势下, 新式移动通信网络的设计需要在可靠性上补足短板, 充分考虑以及配合移动基础软件的发展, 从而共同实现对于未来移动生态架构的真正革新, 支撑并催生新的产业增长点.

## 参考文献

---

- 1 Wang Y, Li J, Huang L, et al. 5G mobile: spectrum broadening to higher-frequency bands to support high data rates. *IEEE Veh Technol Mag*, 2014, 9: 39–46
- 2 Wu G, Wang H M, Wu Q Y. An algorithm on location management and reliable communication for mobile agents. *J Software*, 2002, 13: 269–273 [吴刚, 王怀民, 吴泉源. 一个移动智能体位置管理与可靠通信的算法. *软件学报*, 2002, 13: 269–273]
- 3 赵晨. 5G 网络多场景覆盖策略研究. *信息通信*, 2016, 9: 227–229
- 4 Cai H-Q, Zhang Y, Huang G, et al. Approach to scheduling network requests in Android apps. *J Software*, 2017, 28: 3367–3384 [蔡华谦, 张颖, 黄罡, 等. 一种优化安卓应用 3G/4G 网络请求能耗的方法. *软件学报*, 2017, 28: 3367–3384]
- 5 Vereecken W, van Heddeghem W, Deruyck M, et al. Power consumption in telecommunication networks: overview and reduction strategies. *IEEE Commun Mag*, 2011, 49: 62–69
- 6 Zhou Z, Zhou Y, Xiao J J. Survey on augmented virtual environment and augmented reality. *Sci Sin Inform*, 2015, 45: 157–180 [周忠, 周颐, 肖江剑. 虚拟现实增强技术综述. *中国科学: 信息科学*, 2015, 45: 157–180]
- 7 Cheng J, Chen W, Tao F, et al. Industrial IoT in 5G environment towards smart manufacturing. *J Industrial Inf Integration*, 2018, 10: 10–19
- 8 You X H, Yin H, Wu H Q. On 6G and wide-area IoT. *Chinese J Int Things*, 2020, 4: 3–11 [尤肖虎, 尹浩, 鄂贺铨. 6G 与广域物联网. *物联网学报*, 2020, 4: 3–11]
- 9 Yang Z, He X W, Wu J H, et al. Edge computing technologies for streaming video analytics. *Sci Sin Inform*, 2022, 52: 1–53 [杨铮, 贺晓武, 吴家行, 等. 面向实时视频流分析的边缘计算技术. *中国科学: 信息科学*, 2022. 10.1360/SSI-2021-0133] 2022, 52: 1–53
- 10 Cheng X, Chen C, Zhang W, et al. 5G-enabled cooperative intelligent vehicular (5GenCIV) framework: when Benz meets Marconi. *IEEE Intell Syst*, 2017, 32: 53–59
- 11 Ning H, Wang H, Lin Y, et al. A survey on metaverse: the state-of-the-art, technologies, applications, and challenges. 2021. ArXiv:2111.09673
- 12 Qian Z H, Wang Y J. IoT technology and application. *Acta Electron Sin*, 2012, 40: 1023–1029 [钱志鸿, 王义君. 物联网技术与应用研究. *电子学报*, 2012, 40: 1023–1029]
- 13 Yao F, Kuang L L, Zhan Y F, et al. Key techniques and development trend of antenna arraying for deep space communication. *J Astronaut*, 2010, 31: 2231–2238 [姚飞, 匡麟玲, 詹亚锋, 等. 深空通信天线组阵关键技术及其发展趋势. *宇航学报*, 2010, 10: 2231–2238]
- 14 Liang H, Chen F C, Ji X S, et al. Development status and applied research on mimic technologies for space-ground integration information network. *Sci Sin Inform*, 2019, 49: 799–818 [梁浩, 陈福才, 季新生, 等. 天地一体化信息网络发展与拟态技术应用构想. *中国科学: 信息科学*, 2019, 49: 799–818]
- 15 Dahlman E, Mildh G, Parkvall S, et al. 5G evolution and beyond. *IEICE Trans Commun*, 2021, 104: 984–991
- 16 Zhang P, Tao Y Z, Zhang Z. Survey of several key technologies for 5G. *J Commun*, 2016, 37: 15–29 [张平, 陶运铮, 张治. 5G 若干关键技术评述. *通信学报*, 2016, 37: 15–29]
- 17 You X H, Pan Z W, Gao X Q, et al. The 5G mobile communication: the development trends and its emerging key techniques. *Sci Sin Inform*, 2014, 44: 551–563 [尤肖虎, 潘志文, 高西奇, 等. 5G 移动通信发展趋势与若干关键技术. *中国科学: 信息科学*, 2014, 44: 551–563]
- 18 Luo Z, Fu S, Theis M, et al. Democratizing cellular access with CellBricks. In: *Proceedings of the ACM Special Interest Group on Data Communication*, 2021. 626–640
- 19 Li Y, Li Q, Zhang Z, et al. Beyond 5G: reliable extreme mobility management. In: *Proceedings of the ACM Special Interest Group on Data Communication*, 2020. 344–358
- 20 Xu D, Zhou A, Zhang X, et al. Understanding operational 5G: a first measurement study on its coverage, performance and energy consumption. In: *Proceedings of the ACM Special Interest Group on Data Communication*, 2020. 479–494

- 21 Narayanan A, Zhang X, Zhu R, et al. A variegated look at 5G in the wild: performance, power, and QoE implications. In: Proceedings of the ACM Special Interest Group on Data Communication, 2021. 610–625
- 22 Deng H, Peng C, Fida A, et al. Mobility support in cellular networks: a measurement study on its configurations and implications. In: Proceedings of the ACM Internet Measurement Conference, 2018. 147–160
- 23 Li Y, Lin H, Li Z, et al. A nationwide study on cellular reliability: measurement, analysis, and enhancements. In: Proceedings of the ACM Special Interest Group on Data Communication, 2021. 597–609
- 24 Zhang G, Quek T Q S, Huang A, et al. Delay and reliability tradeoffs in heterogeneous cellular networks. *IEEE Trans Wireless Commun*, 2015, 15: 1101–1113
- 25 Xu T, Zhu C G. Research on ping-pong handover solution in LTE wireless network optimization. *Inform Commun*, 2018, 8: 171–173 [徐彤, 朱长国. LTE 无线网络优化中乒乓切换解决方案的研究. *信息通信*, 2018, 8: 171–173]
- 26 Li Y, Lin H, Li Z, et al. Diagnosis and fixing code for cellular failures at scale. 2021. <https://cellularreliability.github.io/>
- 27 Li Y, Cai L. UAV-assisted dynamic coverage in a heterogeneous cellular system. *IEEE Network*, 2017, 31: 56–61
- 28 3GPP. 3GPP — the 5G standard. 2022. <https://www.3gpp.org/>
- 29 3GPP. Release-15. 2022. <https://www.3gpp.org/release-15/>
- 30 Li Y, Kim K H, Vlachou C, et al. Bridging the data charging gap in the cellular edge. In: Proceedings of the ACM Special Interest Group on Data Communication, 2019. 15–28
- 31 Ahmad M, Jafri S U, Ikram A, et al. A low latency and consistent cellular control plane. In: Proceedings of the ACM Special Interest Group on Data Communication, 2020. 648–661
- 32 Lee J, Lee J, Im Y, et al. CASTLE over the air: distributed scheduling for cellular data transmissions. In: Proceedings of the Annual International Conference on Mobile Systems, Applications, and Services, 2019. 417–429
- 33 Alsaeedy A A R, Chong E K P. A review of mobility management entity in LTE networks: Power consumption and signaling overhead. *Int J Network Mgmt*, 2020, 30: e2088
- 34 Huang Z, Lu J H, Yang S Z. Rapid acquisition of Doppler shift in satellite communications. *Acta Electron Sin*, 2003, 31: 1052–1056 [黄振, 陆建华, 杨士中. 卫星通信中多普勒频移的快速捕获. *电子学报*, 2003, 31: 1052–1056]
- 35 Liao X C, Qiu M, Mai H R. The research and simulation of Doppler effect in wireless sensors networks. *Chinese J Sensors Actuat*, 2007, 12: 2674–2678 [廖惜春, 丘敏, 麦汉荣. 无线传感网络中多普勒效应的研究与仿真. *传感技术学报*, 2007, 12: 2674–2678]
- 36 Zöchmann E, Caban S, Lerch M, et al. Resolving the angular profile of 60 GHz wireless channels by delay-Doppler measurements. In: Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop, 2016. 1–5
- 37 Hadani R, Rakib S, Tsatsanis M, et al. Orthogonal time frequency space modulation. In: Proceedings of the IEEE Wireless Communications and Networking Conference, 2017. 1–6
- 38 Deng H, Li Q, Huang J, et al. iCellSpeed: increasing cellular data speed with device-assisted cell selection. In: Proceedings of the Annual International Conference on Mobile Computing and Networking, 2020. 1–13
- 39 Nguyen B, Zhang T, Radunovic B, et al. ECHO: a reliable distributed cellular core network for hyper-scale public clouds. In: Proceedings of the Annual International Conference on Mobile Computing and Networking, 2018. 163–178
- 40 Li L, Xu K, Li T, et al. A measurement study on multi-path TCP with multiple cellular carriers on high speed rails. In: Proceedings of the ACM Special Interest Group on Data Communication, 2018. 161–175
- 41 Asrese A S, Walegne E A, Bajpai V, et al. Measuring web quality of experience in cellular networks. In: Proceedings of International Conference on Passive and Active Network Measurement, 2019. 18–33
- 42 Zhao P, Zhou Z H. Learning from distribution-changing data streams via decision tree model reuse. *Sci Sin Inform*, 2021, 51: 1–12 [赵鹏, 周志华. 基于决策树模型重用的分布变化流数据学习. *中国科学: 信息科学*, 2021, 51: 1–12]
- 43 Sheoran A, Fahmy S, Osinski M, et al. Experience: towards automated customer issue resolution in cellular networks. In: Proceedings of the Annual International Conference on Mobile Computing and Networking, 2020. 1–13
- 44 Wang J, Zheng Y, Ni Y, et al. An active-passive measurement study of TCP performance over LTE on high-speed rails. In: Proceedings of the Annual International Conference on Mobile Computing and Networking, 2019. 1–16
- 45 Lee J, Lee S, Lee J, et al. PERCEIVE: deep learning-based cellular uplink prediction using real-time scheduling patterns. In: Proceedings of the Annual International Conference on Mobile Systems, Applications, and Services, 2020. 377–390
- 46 Guéraud S, Walsh E K, Cook A E, et al. Validating information during reading: the effect of recency. *J Res Reading*, 2018, 41: 85–101

- 47 Hobfeld T, Varela M, Heegaard P E, et al. Observations on emerging aspects in QoE modeling and their impact on QoE management. In: Proceedings of International Conference on Quality of Multimedia Experience, 2018. 1–6
- 48 Zhang F, Mao J, Liu Y, et al. Cascade or recency: constructing better evaluation metrics for session search. In: Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020. 389–398
- 49 Huawei Smartcare. White Paper for 5G Cloud VR Service Experience Standards. 2019. <https://carrier.huawei.com/~media/CNBGV2/download/products/servies/5G-Cloud-VR.pdf>
- 50 Wang P, Liu B, Hong T. Electric load forecasting with recency effect: A big data approach. *Int J Forecasting*, 2016, 32: 585–597
- 51 高岳, 邵飞, 范鹏贤, 等. 经典塑性力学中加载路径影响的“首因效应”和“近因效应”(英文). *中南大学学报(英文版)*, 2020, 27: 2592–2605
- 52 Srisuresh P, Ford B, Sivakumar S, et al. NAT Behavioral Requirements for ICMP. Technical Report, 2009, RFC 5508
- 53 Kumar A, Postel J, Neuman C, et al. Common DNS Implementation Errors and Suggested Fixes. Technical Report, 1993, RFC 1536
- 54 Chen J S, Yang X J, Huang Z L. The GPRS mobility management. *Telecom World*, 2001, 7: 38–39 [陈金山, 杨学军, 黄载祿. GPRS 的移动性管理. *通讯世界*, 2001, 7: 38–39]
- 55 Zhou Y, Lau B P L, Yuen C, et al. Understanding urban human mobility through crowdsensed data. *IEEE Commun Mag*, 2018, 56: 52–59
- 56 Lan L, Gou X, Mao J, et al. GSM co-channel and adjacent channel interference analysis and optimization. *Tinshhua Sci Technol*, 2011, 16: 583–588
- 57 3GPP. Generation Partnership Project; NR; Technical Specification Group Radio Access Network; Multi-connectivity; Overall description; Stage-2 (Release 15). 3GPP TS 37.340 v15.0.0. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198>
- 58 Polese M, Giordani M, Mezzavilla M, et al. Improved handover through dual connectivity in 5G mmWave mobile networks. *IEEE J Sel Areas Commun*, 2017, 35: 2069–2084
- 59 Agiwal M, Kwon H, Park S, et al. A survey on 4G-5G dual connectivity: road to 5G implementation. *IEEE Access*, 2021, 9: 16193–16210
- 60 Winkler G. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction. Berlin: Springer Science & Business Media, 2012
- 61 Giorno V, Nobile A G. A time-inhomogeneous Prindiville model with failures and repairs. *Mathematics*, 2022, 10: 251
- 62 Otten R H J M, van Ginneken L P P P. The Annealing Algorithm. New York: Springer Science & Business Media, 2012
- 63 Xiao A, Liu Y, Li Y, et al. An in-depth study of commercial MVNO: measurement and optimization. In: Proceedings of the Annual International Conference on Mobile Systems, Applications, and Services, 2019. 457–468
- 64 Li Y, Zheng J, Li Z, et al. Understanding the ecosystem and addressing the fundamental concerns of commercial MVNO. *IEEE ACM Trans Networking*, 2020, 28: 1364–1377
- 65 Zeqiri R, Idrizi F, Halimi H. Comparison of algorithms and technologies 2G, 3G, 4G and 5G. In: Proceedings of International Symposium on Multidisciplinary Studies and Innovative Technologies, 2019. 1–4
- 66 Haile H, Grinnemo K J, Ferlin S, et al. End-to-end congestion control approaches for high throughput and low delay in 4G/5G cellular networks. *Comput Networks*, 2021, 186: 107692
- 67 Jo Y, Kim H, Lim J, et al. Self-optimization of coverage and system throughput in 5G heterogeneous ultra-dense networks. *IEEE Wireless Commun Lett*, 2019, 9: 285–288
- 68 Mei H. New Era, New Mission and New Journey. *Sci Sin Inform*, 2018, 48: 1–2 [梅宏. 新时代新征程新使命. *中国科学: 信息科学*, 2018, 48: 1–2]
- 69 邹琳琳. 5G 低功耗大连接场景下的能效优化研究. 北京: 北京邮电大学, 2019
- 70 陈山枝, 王胡成, 时岩. 5G 移动性管理技术. 北京: 人民邮电出版社, 2019
- 71 Li J, Qiu S B, Li H J, et al. Smart vehicle for smart city. *Sci Sin Inform*, 2016, 46: 551–559 [李骏, 邱少波, 李红建, 等. 智慧城市的智能汽车. *中国科学: 信息科学*, 2016, 46: 551–559]
- 72 Gui W H, Zeng Z H, Chen X F, et al. Knowledge-driven process industry smart manufacturing. *Sci Sin Inform*, 2020, 50: 1345–1360 [桂卫华, 曾朝晖, 陈晓方, 等. 知识驱动的流程工业智能制造. *中国科学: 信息科学*, 2020, 50:

1345–1360]

73 Zhang P, Niu K, Tian H. Technology prospect of 6G mobile communications. *J Commun*, 2019, 40: 141–148 [张平, 牛凯, 田辉, 等. 6G 移动通信技术展望. *通信学报*, 2019, 40: 141–148]

## Recency effect and self-regulating design of mobile cellular systems in the context of interlaced generations: network bandwidth, power efficiency, and connection reliability

Yunhao LIU\*, Zhenhua LI, Yang LI, Hao LIN, Liangyi GONG & Yawen ZHENG

*School of Information Science and Technology, Tsinghua University, Beijing 100084, China*

\* Corresponding author. E-mail: yunhao@tsinghua.edu.cn

**Abstract** The mobile cellular system is undergoing a generation-interlaced evolution from 2G/3G/4G to 5G, with various signals and protocols coexisting for a long period. A crucial but challenging problem that emerges is how to ensure high performance (fast transmission), high energy efficiency (low power consumption), and high reliability (stable connection) for the services in a large, heterogeneous, and highly dynamic cellular environment. This problem simultaneously involves mobile phone manufacturers, network service providers, and OS/application developers. In this paper, we comprehensively review the representative related works in recent years, and learn that a series of performance anomalies in cellular systems is in fact attributed to the “recency effect” at the core mechanisms. Upon selecting which mobile network to use, there exist generation biases and unconscious performance-first tendencies affected by public opinions and commercial promotions. The biases and tendencies lead to knock-on effects like the “loss-making migration” towards 5G, the “saturated hunger” of network capacity, and the “ping-pong jumping” among base station connections. To overcome the above shortcomings, we propose self-regulating techniques based on 4G/5G dual-connectivity and Time-Inhomogenous Markov Process, which reduce the failure rate of 5G devices by 40% and the average failure duration of all devices by 36%. We also discuss other insights and innovations based on large-scale industrial practice.

**Keywords** interlaced generations, 5G, recency effect, loss-making migration, self-regulating design