

# 云存储价格战背后的科研缺失

李振华 李 健  
清华大学

关键词：云存储 网盘 价格战

## 云存储的迅速繁荣

最近几年，云计算技术的迅猛发展和互联网带宽的多次提速使得电子数据的云端存储与普通访问成为可能。云存储服务（俗称“网盘”）让互联网用户能够方便地存取和分享数据——不管用户在何时何地、使用什么终端设备（个人PC、平板电脑或智能手机），在云端存放的数据，如文档、图片、音乐或视频等都能自动同步到该用户所有的在线设备，并可以与其他用户分享。

国外有代表性的云存储服务包括 Dropbox、微软的 OneDrive、谷歌的 Google Drive、苹果的 iCloud Drive、SugarSync 和 Ubuntu One 等。国内主流的云存储服务有金山快盘、360 云盘、百度网盘、115 网盘、新浪微盘和阿里酷盘等。这些云存储服务中，仅 Dropbox 就拥有超过 3 亿的用户，日均新增和更新文件达数十亿个<sup>[1,2]</sup>。OneDrive、Google Drive、360 云盘等每家都拥有过亿用户。上述统计还不包括依赖于云存储服务的一系列

衍生服务，比如谷歌公司著名的 Gmail 邮箱。Google Docs 在线编辑系统已经与 Google Drive 无缝整合，微软的 OneDrive 已经和 Microsoft Word、Microsoft

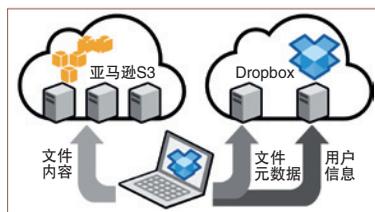


图1 Dropbox系统架构

PowerPoint 一样成为 Microsoft Office 的一个基本构件，而国内著名的小米手机云服务则部分依赖于金山快盘。

## 范例：来自Dropbox的启发

2008年9月，Dropbox 发布第一个公开版本，宣告商业云存储服务的诞生。直到今天，Dropbox 依然是云存储行业的技术和市场领导者，可谓“一直被模仿，从未被超越”。那么，Dropbox 的优势是什么呢？

**系统架构** 2006年3月，亚马逊推出 S3 (Simple Storage Service)，这是世界上第一个商业云存储基础设施，提供极为简单却极具扩展性的 RESTful（自表义、无状态的）访问接口。然而，普通用户要使用 S3 是极为困难的，因为它连基本的图形界面和“文件（夹）”的概念都没有。2007年，Dropbox 公司一成立，就敏锐地抓住了 S3 的后端优势和前端不足，在 S3 和用户之间迅速架起了云存储服务的桥梁。如图 1 所示，Dropbox 自行搭建了一个规模较小的私有云，以维护那些重要的、敏感的元数据和用户信息<sup>[3]</sup>。同时，Dropbox 充分利用 S3 的云计算优势，轻松地应对了后来用户量和数据规模的多次“井喷”。

**技术功能** 受益于 Dropbox 新颖而独特的系统架构，研发人员几乎不需要关心基础设施及其扩展性、可用性和可靠性，节省了巨大的刚性开支。因此，Dropbox 的研发资源几乎全部投入到各项领先而创新的技术功能中，包括小文件批量同步、文件修改增量同步、协同文档编

表1 国内外代表性云存储服务的技术功能和市场运营策略对比  
(2014年6月, 个人PC客户端)

服务	技术功能					市场运营			
	小文件 批量 同步	文件修改 增量同步	协同文 档编辑	文件压缩 和消重	文件版 本回滚	初始 空间	付费空间	跨平台 (特别是Linux)	应用聚合
Dropbox	√	√	√	√	√	2 GB	每月最低9.99美元	√	第三方应用
OneDrive	-	-	√	-	-	7 GB	每月最低2.08美元	-	微软Office
Google Drive	-	-	√	-	√	15 GB	每月最低2美元	-	Google Docs
iCloud Drive	-	-	-	-	-	5 GB	每月最低1美元	-	iCloud
金山快盘	√	-	√	消重, 不压缩	√	2 TB	-	√	小米云服务
360云盘	√	√	-	√	√	36 TB	-	-	360安全卫士
百度网盘	√	-	-	消重, 不压缩	√	2 TB	-	-	离线下载

辑、文件压缩和消重、文件版本回滚等。仅这5项技术功能就为Dropbox构建了坚固的技术壁垒。

**市场运营** Dropbox自上线6年来一直只给免费用户2GB的初始空间,但充分鼓励免费用户通过各种积极的方式获得更多存储空间,如邀请朋友、加入社区、共享文件、活跃使用等。看似吝啬的初始空间和广告策略(Dropbox几乎从来不做花钱的广告),却强化了Dropbox利用社会网络进行市场营销的实际效果。今天,虽然Dropbox的注册用户数可能略低于由互联网巨头支撑的OneDrive, Google Drive和iCloud Drive,但其用户活跃度和盈利状况却是其他云存储服务提供商望尘莫及的。

## 价格战：技术和市场的博弈

早在2012年,华尔街对Dropbox的估值就已经接近100

亿美元,这给互联网巨头们带来了极大的压力和动力。各大互联网公司纷纷投入巨资加入云存储服务市场,各显身手、激烈竞争。之后的两年里,我们见证了美国和中国云存储服务提供商在应对技术和市场博弈时所采取的迥异策略,如表1所示。

**美国云存储巨头：降低付费门槛,聚合强势应用,力图黏住用户。**与Dropbox相比,微软、谷歌、苹果公司都提高了云存储服务的初始空间(2GB → 7/15/5 GB),同时降低了付费空间的门槛(每月9.99美元 → 每月2/2/1美元),以吸引更多用户尝试和体验。在应用聚合方面,Dropbox主要是间接聚合了一系列第三方相对弱势的应用,而微软、谷歌、苹果则直接聚合本公司的强势应用,力图强强联合、黏住用户。

**中国云存储巨头：完全免费,亏本运营,牺牲部分功能和性能。**为了在最短的时间内

争取到最多的用户,金山、百度、360、115等公司将云存储价格大战推到违反商业逻辑的极限。他们相继宣布各自的云存储服务免费,并且初始空间大得惊人——金山和百度提供2TB,360赠送36TB,115网盘则宣布“无限空间,每时每刻自动扩容”。从表面上看,优惠幅度很大,但用户不禁对这些国内云存储巨头产生疑惑:他们有什么?真的具有足够的基础设施和技术能力来为用户提供满意的服务吗?他们图什么?亏本运营除了能赚到虚高的用户数目,还能赚到什么?

**云存储小公司：强制付费,或关门大吉。**前有Dropbox的技术壁垒,后有互联网巨头的价格大战,面对竞争激烈的云存储市场,诸多创业型的小公司纷纷选择退让或者放弃。曾经名噪一时的SugarSync,在技术上并不逊色于Dropbox,但从2014年初,SugarSync开始删除免费用户账号及其数据,以保留忠实用户、

降低基础设施开支。同样著名的云存储服务 Ubuntu One 的提供商,也在 2014 年 5 月直接关闭了整个系统。对于这些没有强大后台支撑、不具备价格战能力的云存储小公司来说,在市场红海中急流勇退,也不失为明智之举。

我们在此最关心的是:在云存储技术和市场的博弈中,中国的主流云存储服务同美国相比,为何出现如此大的反差?其根源何在?

从表面上看,云存储服务在中国看似“天上掉馅饼”的价格大战里被严重扭曲,由“红海”变为“血海”。实际上,对于绝大多数人来说,即使把他拥有的几个计算设备加在一起,存储空间也达不到 2TB。因此,动辄宣称免费赠送数 TB 空间的网盘,对用户来说并无实质意义。而更大的隐患是:由于存储服务完全免费,系统亏本运营,母公司不可能向云存储服务投入太多研发资源。在技术和功能落后的情况下,服务性能差强人意,服务功能存在关键缺失,即使空间大得惊人,又有多少实际价值呢?

中国的主流云存储服务同美国相比出现巨大反差的根源是,不管对于云服务的提供商还是用户来说,云计算的思维模式在中国还没有被真正、深度地接受。中国的信息技术领域整体还处在“重硬件、重指标”的刚性思维阶段,对云计算这样一个“拿来主义”的新生事物,缺乏“重技术、重服务”

的柔性思维模式。

把这个问题放大来看,可以追溯到当前特殊的国情。十多年来,互联网构造的扁平化虚拟社会(如腾讯微信、新浪微博、网易新闻等)与几千年传统文化形成的垂直层次化的现实社会之间的内在矛盾,造成了当代中国社会(对数量指标)的普遍浮躁。即便是向来标榜高科技、开放性的中国互联网领域,也未能幸免。

## 深入:360云盘技术团队

为了获得中国云存储产业界的第一手资料,我们通过 360 搜索和 360 系统部门的朋友引荐,拜访了 360 云盘技术团队,并和他们就云存储技术和市场问题进行了深入交流。

令我们佩服的是:整个 360 云盘技术团队仅有 3 个人,其中两个人还是“兼职的”——他们同时归属于其他产品团队。虽然 360 云盘系统的研究是建立在 360 公司多年来先进的基础设施和丰富的部署经验积累之上,但 3 个人管理数千台服务器、服务上千万用户,仍然让人感觉不可思议。

但令人遗憾的是:尽管我们向其技术团队汇报了实际测量到的 360 云盘服务的诸多问题,提出了解决方案,而且也得到了他们的认可,但是,由于受限于中国云存储产业界特殊的发展模式和竞争格局,360 云盘至今也未

能解决这些问题。

## 科研探索:“节流盒子”与“清云网盘”

Dropbox 的 5 项技术功能为其构造了坚固的技术壁垒,不仅让行业竞争者望尘莫及,甚至连学术界都知之甚少,还停留在表象层面的“黑盒”研究<sup>[4-7]</sup>。对于 Dropbox,我们也从多个层面开展了一些研究,并将它们总结为“节流盒子”(SavingBox)<sup>[8]</sup>和“清云网盘”(THUCloud Disk)<sup>[9]</sup>。

### 节流盒子

由于超过十亿互联网用户已经将他们的数据托管在云存储中,因此云存储服务如何有效地利用网络资源进行经济高效的数据同步就变得极为重要。Dropbox 类云存储服务通常采用两个方法来降低网络流量开销。第一个方法,当数据更新时,客户端计算更新文件(相对云端文件)的二值差异,并且只将此二值差异(即文件中被改变的那部分)发送到云端。第二个方法,在任何更新发送到云端之前,客户端都会进行压缩。例如,当我们向 Dropbox 同步文件夹中一个现存文件添加 100 MB 的相同字符时,二值差异的大小为 100MB,此次文件修改只消耗了约 40KB 的网络流量——这和仅添加一个字符所消耗的网络流量(约 38KB)很接近。

虽然如此,当面临频繁短促

的数据更新（如协同文档编辑、系统日志追加、数据库更新）时，云存储服务所产生的网络流量却呈现出“病态的低效”。具体的表现是，每当同步文件夹中发生数据更新时，云存储服务当前普遍采用的“更新触发的实时同步机制”（Update-triggered Real-time Synchronization, URS）就会立刻计算二值差异并进行压缩，再将压缩后的二值差异发送到云端。这个过程不可避免地要同时发送一系列会话维护数据，包括域名服务器（DNS）查询、传输控制协议 / 安全超文本传输协议（TCP/HTTPS）连接的建立和维护、数据索引等操作所需要的数据。所以，当用户数据流呈现频繁短促的数据更新模式时，会话维护流量将远远超出实际更新内容的数据流量，从而给云端、客户端以及传输网络带来沉重而非必要的流量负担，我们称之为云存储服务的“同步流量滥用问题”。从本质上讲，该问题源自更新触发的实时同步机制的更新敏感性。

为了深入理解流量滥用问题，我们对 Dropbox 进行了测量，人为地生成一系列数据更新模式来观察相应的同步网络流量。虽然 Dropbox 是一个闭源商业应用，并且它的绝大部分通信数据包使用加密套接字协议层（Security Socket Layer, SSL）加密，但我们依然能够使用 Wireshark 网络协议分析器，通过“黑盒测试”来理解 Dropbox 的工作原理。配合分析 Dropbox 数据包的时序，我

们量化地解释了为什么在处理频繁短促的数据更新时，会话维护流量远远超出实际数据流量。更重要的是，我们发现了触发 Dropbox URS 同步机制的本地操作系统的特性，将 Dropbox 同步的整个过程分解为一系列“原子

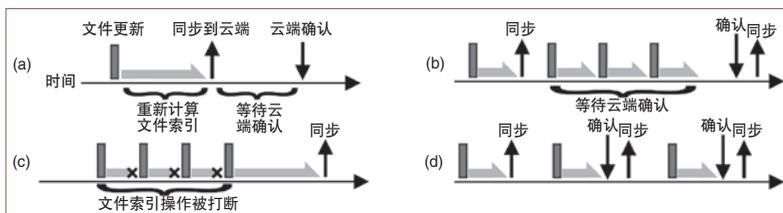


图2 Dropbox数据同步的四个“原子过程”

过程”并加以探究，如图 2 所示。

基于对 Dropbox 数据同步原理的深度理解，我们设计了一个名为“节流盒子”的中间件，以解决 Dropbox 类云存储服务的流量滥用问题，其本质是“批量延迟的同步机制”（Update-batched Delayed Synchronization, UDS）。如图 3 所示，节流盒子介于用户文件系统和云存储应用之间，独立于任何特定的云存储系统，并

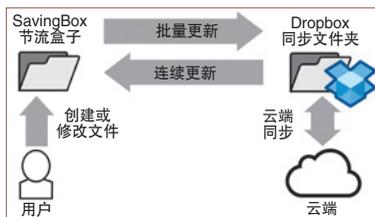


图3 节流盒子和UDS高效批量同步算法

且不需要对原始云存储系统进行任何改变，简单易用。具体地说，首先构建一个名为“节流盒子”的文件夹以替代原始的同步文件夹的功能。一旦探测到频繁短促

的数据更新，就适当延迟同步，将多次小更新打包成一次大更新同步到云端。节流盒子强迫云存储应用合并那些可能导致流量滥用问题的数据更新。在实际环境下，所导致的额外同步延迟仅为数秒，一般用户感觉不到。

## 清云网盘

为了广泛、深入地研究云存储服务，我们搭建了“清云网盘”系统，试图跨越云存储服务的所有技术层面，系统化地研究云存储服务的体系结构、基础设施、数据模型、访问接口、元数据管理及数据同步协议。

在搭建“清云网盘”的过程中，我们特别关注两个方面：(1) 主流性，即整个系统的体系结构能够代表当前工业界主流云存储系统，特别是 Dropbox 的模式和趋势；(2) 开放性，即所有框架和代码均使用开源软件构建，使其他研究者能够读懂、修改。

为保证“清云网盘”的主流性，我们效仿 Dropbox，采用了“数据云”和“控制云”相分离的体系结构——数据云存储实质数据内容，控制云维护数据索引和用户状态信息，客户端同时与数据云和控制云交互。分离式体系结



图4 “清云网盘”网页版截图

构既有利于数据的可靠存储与快速扩展，又便于控制逻辑的独立开发与灵活调试。

为保证“清云网盘”的开放性，我们使用了 Ubuntu 开源操作系统、Openstack 开源云计算框架（特别是其中的 Swift 和 Keystone 组件）、Django 和 WordPress 开源网站框架、MySQL 开源数据库，以及其他一系列经典开源软件，包括 Memcached、OpenSSH、Python、PHP、Apache 和 Rsync 等。

2014年，在经过大量局域网测试之后，我们把整个清云网盘系统（包括硬件服务器和交换机）从清华大学软件学院搬迁到教育网（CERNET）数据中心，并购买了专门的域名（www.thucloud.com）和公网地址，使之成为一个名副其实的互联网云存储服务，如图4所示。“清云网盘”系统目前已经稳定运行半年多。

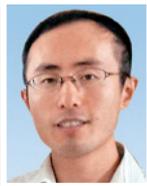
受限于高校的资源平台，“清云网盘”团队既不追求注册用户数，也不追求盈利，只在乎系统本身的科技含量，并努力把我们的技术方案模块化和普适化，同时形成清晰完备的技术文档，期望为“中国云科技发展‘十二五’专项规划”作出贡献。

## 结语

叔本华精辟地分析过人类痛苦的来源：“这个世界最大的痛苦不在于不圆满，而在于被扭曲”。我们有充足的证据可以断言，中国今天的云存储行业确实被价格大战所扭曲，因而在一定程度上忽视了

用户真正关心的功能和性能，对用户的黏性相对薄弱。我们认为问题的解决之道在于回归本源，效仿云存储服务的先驱和典范 Dropbox——只要真正从用户出发，真心为用户着想（甚至有时候只是为了一群小众用户），那么，即使它只提供 2GB 的初始空间，即使它的数据中心远在太平洋彼岸，即使它经常由于莫名其妙的原因被屏蔽封杀，用户也依然对它无限忠诚、不离不弃。■

**致谢：**本文工作得到了中国博士后科学基金（2014M550735）“面向普适云存储服务的高效能数据同步协议研究”的资助。



李振华

CCF会员。清华大学博士后。主要研究方向为云计算、云存储、互联网、智能路由。  
lizhenhua1983@tsinghua.edu.cn



李健

CCF会员。清华大学博士后。主要研究方向为云计算、云存储、互联网、流媒体。  
upton@tsinghua.edu.cn

## 参考文献

- [1] Dropbox reaches 300 million users, adding 100 million users in just six months. <http://thenextweb.com/insider/2014/05/29/dropbox-reaches-300m-users-adding-100m-users-just-six-months>.
- [2] Dropbox clears 1 billion file uploads per day. <http://www.cnet.com/news/dropbox-clears-1-billion-file-uploads-per-day/>.
- [3] Z. Li, C. Wilson, Z. Jiang, et al.. Efficient batched synchronization in Dropbox-like cloud storage services. Middleware, 2013.
- [4] W. Hu, T. Yang, and J.N. Matthews. The good, the bad and the ugly of consumer cloud storage. *ACM SIGOPS Operating Systems Review*, 2010; 44(3):110~115.
- [5] I. Drago, M. Mellia, M.M. Munaf o, et al.. Inside Dropbox:

- understanding personal cloud storage services. IMC, 2012.
- [6] I. Drago, E. Bocchi, M. Mellia, et al., Benchmarking personal cloud storage. IMC, 2013.
- [7] D. Kholia and P. Wegrzyn. Looking inside the (Drop) box. The 7th USENIX Workshop on Offensive Technologies (WOOT), 2013.
- [8] Y. Zhang, C. Dragga, A. Arpaci-Dusseau, et al.. ViewBox: integrating local file systems with cloud storage services. FAST, 2014.
- [9] Z. Li, H. Xiao, L. Cheng, et al.. T-CloudDisk: A tunable cloud storage service for flexible batched synchronization. Middleware (Live Demo), 2013.