★CACM 中国版★

计算机协会通讯

CACM.ACM.ORG

2014 年第 57 卷第 04 期

增强现实系统的安全和隐私

多核硬件

速率限制状态

大规模在线公开课程中的人与事

有界双调和权重

Association for Computing Machinery

ACM 通讯

精选文章中文译文

ACM中国理事会办公室

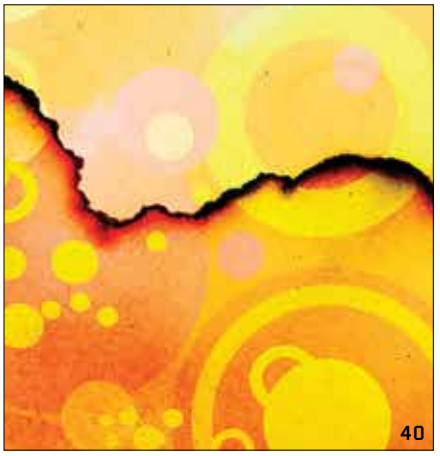
中国北京清华大学
东主楼 11-236 室
邮编：100084
电话：+86-10-62785025
电子邮件：acmchina@acm.org
联系人：辛爽

ACM通讯

(ISSN 0001-0782) 由计算机协会（2 Penn Plaza, Suite 701, New York, NY 10121-0701）按月发行。

Association for Computing Machinery

观点

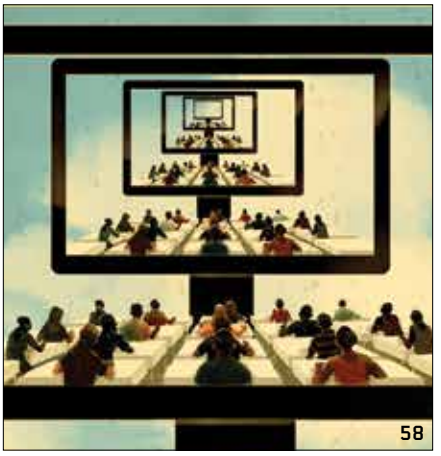


实践

封面故事：
通过耳麦 / 眼镜、智能手机和其他移动设备实现的增强现实应用在创造精彩与裨益的同时，也带来了切实的安全和隐私问题。本月的封面故事（第 88 页）探讨了各种风险，并认为目前此项技术尚属早期，正是解决这些问题的良机。封面照片由 Michael Zhang 提供。

图片由 SHUTTERSTOCK.COM 提供；插图由 ANTHONY FREDA、BARRY DOWNARD 提供

投稿文章



评论文章

研究亮点

Association for Computing Machinery
Advancing Computing as a Science & Profession

DOI:10.1145/2580945

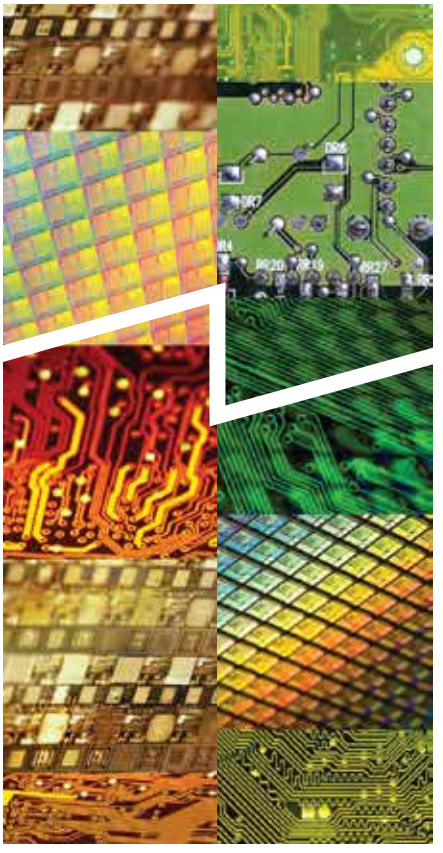
Uzi Vishkin

观点
通用并行处理的多核硬件已遇困境?

当前一代通用多核硬件必须得到改进，以期支持更多应用领域并可进行符合成本效益的并行编程。

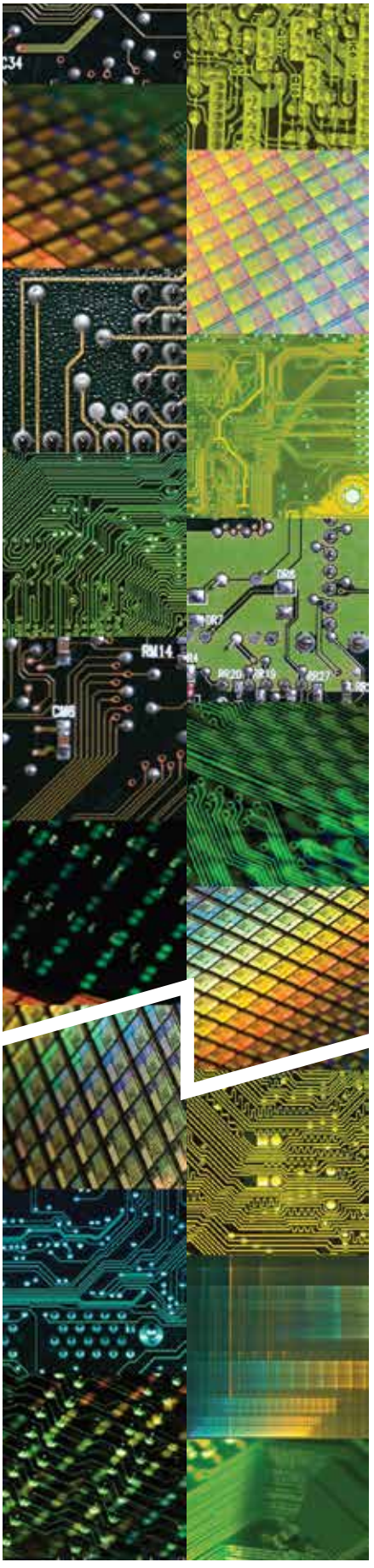
最近10年，主流通用计算机性能增长大多依靠增加处理器核心数量。总体来说，并行计算毫无疑问取得了巨大进步。在Google、Facebook等公司，并行计算以GPU的形式在超级计算应用中得到了规模空前的运用。然而，本文不讨论这些辉煌成就。并行计算今后要想追求进步，必须从审视目前的一些缺点开始，这正是本文的目标。本文希望就如何最好地弥补这些缺点，走向并行计算的黄金时代引发一场建设性的讨论。

如今的并行体系结构能够在规则程序（例如稠密矩阵型）上获得较高的加速比。然而，对于*其他通常所谓的“不规则”程序，或当寻求“强可扩展性”时，这些体系结构大多能力不足*。强可扩展性是指在固定问题的输入规模时，将处理器核心数量的增加转化为更短的执行时间的能力。通常，只有问题的现有算法能够被映射到高度并行和严格结构化的程序，才有可能比最快的串行算法获得好的加速比。但是，即便对于规则并行编程，符合成本效益仍然是个问题。程序员实现基本加速比所需的工作量比基本的串行编程要大得多（在领域特定语言存在有限的例外，即获得基本加速比所需的工作量可以在一定程度上减少）。值得注意的是，过去十年间高端通用桌面应用程序的创新微乎其微（或许计算机图形学是个例外），相比移动应用程序和互联网应用程序，这一点尤为突出。此外，英特尔等厂商2005年预测[2]，2013年的主流处理器将拥有数百个核心（“转向多核时代的范式转变”），但现实却是依然只有少数几个核心。这或许是因为这段时间内通用台式机厂商之间的竞争减弱，它们没有动力冒这种范式转变的风险。但是，这是否意味着这些计算机的硬件遭遇困境?



我认为答案是肯定的。还有更多应用领域可以大大受益于利用不规则并行性提高加速比。以下是部分这些领域的列表。

- *生物信息学*。基因组学涉及到许多图和其他组合问题，这些问题很少是规则的。
- *计算机视觉*。图形处理单元(GPU)只能较好地支持OpenCV库的一小部分。其他OpenCV原语通常是不规则的。



- *科学计算*。稀疏矩阵运算和需要运行时自适应的问题，例如解线性方程组的多尺度法，或量子系统的动力学模型 。
- 数据压缩。
- 信号处理中的稀疏感知与恢复。
- 面向设计和模拟的电子设计自动化 (EDA)。
- 数据同化；
- 图模型的众多用途：
  - 数据分析。
  - 社交网络分析。
  - 流行病网络。
  - 置信传播。
- *开放性问题和编程*。该领域的特征体现在提出问题的方式和开发计算机程序的方式，而不是特定应用。许多问题的初始理解并不意味着明确的输出，有时甚至输入、定义本身都适合不规则编程。

如今的通用多核硬件没有为上述任何领域提供充分的支持。这种硬件必须得到改进，以期支持更多领域，并且可进行符合成本效益的并行编程。这一改进需要改变当前的硬件及包含硬件在内的整个生态系统，包括编程实践和编译器。此类系统的主要问题是：程序员应该如何在实现的算法中表达并行性；编译器和硬件 / 软件运行时系统的责任是什么；硬件将如何尽可能有效执行最终可运行的并行性。我认为，对于不规则问题，程序员所能发现的并行性（知道哪些运算能同时执行）远超当前和不久以后的编译器技术能从串行代码中提取的并行性，这类并行性大部分是细粒度的，并且可弱（能同时执行的运算非常少）可强（许多此类运算），即使在同一并行算法的不同步骤中也是如此。因此，系统设计人员必须“时刻留心”，将程序员提供的并行性视为*最宝贵的资源*。特别是：鼓励程序员表达其发现的所有并行性（当然，在用编程语言支持这种表达后）并优化编译器、运行时系统和硬件，从而从程序员提供的任何并行性中获得可能的最佳性能。尤其是最大限度减少将程序中的的任何并行性转化为性能的开销。

可以采取的措施包括共享（片上）缓存、低开销控制机制、高带宽、低延迟互连网络和灵活的数据获取和存储机制。很多人自然会对该列表感到疑惑并问：我们不是知道这些方法了吗？这些方法不是已经实现了吗？经过仔细研究，你会发现在当前计算机系统的设计中，在采用这些方法的时候由于考虑了与不规则并行性能相竞争的其它目标（未能集中目标），使得这些方法的效果不明显，导致目前并行体系结构和不规则并行算法之间产生了鸿沟。以下是制约多核硬件在处理不规则并行程序时的性能的几个竞争目标示例。

竞争目标：*提高吞吐量*。目前多处理器由多个紧密耦合的处理器组成，这些处理器的协调和使用由单个操作系统控制。当体系结构的目标是最大化在给定时间里完成的（可能是较短的）作业数量时，这是一种常见的方法。（有时该目标被称为优化吞吐量。）作为一个软件系统，操作系统在线程管理方面（包括线程初始化、将线程动态分配给硬件、线程终止）会产生不必要的高开销。然而，如果主要目标放在对不规则并行程序的低开销支持上，我们本该看到这些功能更多地向硬件迁移，*细粒度程序更是如此，而许多不规则程序都是细粒度程序*。

竞争目标：*最大化峰值性能*。GPU 设计的引导方向似乎是在一块硅晶片尽量多安装功能单元，试

图最大限度地提高峰值性能（例如 FLOPS）。但是，有效支持不规则程序、强可扩展性和更强的持续性能（而不是峰值性能）是一个不同的目标，它需要在硬件和程序员两方面作出不同的选择。硬件方面的例子是：通常不能假设功能单元需要处理的数据可在功能单元附近提供，也不能假设在使用向量功能单元的时候，数据可以以高速率和结构化的方式提供给功能单元。我们也不能期望程序员用简单并行编程解决此类不规则数据问题。有很多例子表明，GPU 需要程序员提供严格结构化的数据。

竞争目标：*局部性最大化。*高速缓存在串行和并行体系结构中所发挥的相应作用至少有助于解释部分问题。

*串行体系结构中的高速缓存。*并行计算始于成功的通用编程模型。20 世纪 80 年代期间，内存延迟的降低开始跟不上串行处理器速度的提升，高速缓存成为了继续支持串行编程模型的解决方案。研究发现，串行程序倾向于重复利用数据（或最近使用过的数据附近的地址），这一规律也称为“局部性原则”，这意味着高速缓存通常可以继续支持该模型，从而缓解问题。因此，尽管局部性成为了优化串行硬件的重要主题，但人们不允许它影响日常编程的基本编程模型。针对局部性的编程工作，也只是由相对很少的“性能程序员”完成的。

*本地并行内存。*并行计算从没拥有真正成功的通用编程模型，因此人们没有足够的动力来投入精力继续支持该模型。从早期开始，并行处理体系结构就由多个处理器耦合而成，其中每个处理器都带有容量相当大的本地存储单元。之前提

并行计算今后要想追求进步，必须从审视目前的一些缺点开始。

到过，关键原因之一是追求更高的峰值性能 (FLOPS)。这意味着，在资金预算、硅片面积预算或功率预算一定的情况下，最大限度增加功能单元及其相邻内存的数量。我还注意到，人们似乎不会优先考虑为了提升持续性能（也许以牺牲峰值性能为代价）而牺牲其中一些预算。因此，将并行任务映射到这些本地内存中已经成为程序员的巨大责任，也成为了将并行计算扩展到规则应用之外和简化并行编程的一个主要障碍。

竞争目标：*优先处理高度并行应用。*当前的设计似乎期望应用具有非常高的并行性。我认为，一般来说，该观点过于乐观。串行计算机体系结构给我们的一个教训是，需要让任何通用硬件平台经过重要的基准压力测试。对不规则算法和问题而言，在目前的并行机器上的并行性规模缩减往往也是一个问题。例如，对于图的广度优先搜索，某些问题实例（例如，随机图）可能会提供大量的并行性，而另一些不会（例如，大直径图）。在一些算法中，算法的不同步骤并行性迥异。例如，标准最大流算法提供了一项实用的压力测试。这些最大流算法对直径不断增大的图反复进行广度优先搜索，因此并行性随着算法的进度而降低，只能处理高并行性的体系结构在此失效。

竞争目标：*节能优先于程序员的生产率。*登纳德缩放比例定律 (Dennard Scaling) 临近极限，计算机功耗改进逐渐放缓，这些都构成了重大问题。[5] 功耗比程序员的生产率更容易量化，也更接近硬件设计人员的舒适区。这可以解释我们经常听到的一个观点，即并行程序员必须主动承担为降低功耗而编程的责任，这种观点影响了一些设计决策。我发现这一趋势有两个问题，一个相当具体，另一个则是更原则性的。具体的问题是，不规则问题使程序员难以（如果可能）符合这些设计决策。一般性问题是，这一基本观点似乎“违背历史”。工业革命的大部分进步归功于使用更多动力来减少人的劳动。计算性能今后的进步能建立在逆转这一趋势的基础上吗？

读者需要了解，本文中质疑厂商硬件的方法还远未获得一致赞同。实际上，许多作者都试图迎合此类硬件，对这些硬件带来的限制进行建模，以在算法设计中满足这些限制。大容量同步并行 (BSP)[6] 和最近 Ballard 等人 [1] 提出的多种通信回避算法是规则算法方面巨大成就的显著例子。然而，最新现状依然是：除非问题实例能够被映射到稠密矩阵结构，否则就得不到有效解决。经过多年的并行算法研究，我认为如此根本性的变化在现实中是不可行的。

有趣的是，这种通信避免学派观点和本文观点之间的差异并没有表面上那么大。在解释原因之前，我要指出，在大型科学应用推动下，并行计算方面的政府投资历来大多与大型高性能计算机的前沿技术相关。在本文中，我主张发展小规模

系统中的并行计算，以期改善编程和通用应用的运行时、易用性和灵活性。如果能够成功，此类小规模系统反过来也能够为大型计算机提供更好的组成部件（例如，节点）。例如，Edwards 和 Vishkin 最近发表的文章[3]指出，在仍然使用相同互连硬件的情况下，大型计算机的这种组织能够让每对节点之间的有效带宽增加一倍。许多大型计算机的运行方式是从一个节点向另一个节点发送消息。为了更好地利用带宽，发送节点在发送消息之前用数据压缩算法对其进行压缩，接收节点则应用与之匹配的解压缩算法进行解压缩。在节点上使用 XMT 型计算机（参见后面的参考文献）通常能够使压缩消息的大小减半，同时不会导致压缩和解压缩的延迟增加。就通信回避学派观点而言，该方向提供了一个有趣的中间地带。即，在大型计算机节点内使用高带宽且易于编程的并行计算范式，在节点之间使用通信回避范式。

我教授高年级计算工程专业课程。在此之前，学生们必须学完几乎所有必需的编程、数据结构和算法课程，许多学生还通过实习和研究获得了一些软件开发和应用方面的经验。听取他们的意见对我来说很有趣，因为他们已经知道许多，但尚未承受很多循规蹈矩的压力，例如来自雇主的压力。他们发现难以接受并行编程的现状。在我向学生们讲述规则程序和不规则程序间的悬殊差别时，他们指出，自己遇到过的绝大多数程序都是不规则的。可以理解，他们对并行编程不能追求同样灵活的风格有所不满。对他们而言，*标准的计算机课程和实践培养了他们“亲自动手”解决问题的方式*，如今并行硬件要求的并行编程类型正在削弱这种方式。这关系到前文关于开放性问题和编程的讨论。为“不明确”的问题开发程序，不同程序员的开发方式迥然不同，每个人开发出来的程序也是如此，在反映计算机科学的“灵魂”方面，个性和创造力不亚于易于理解的应用，特别是，个性和创造力对吸引人才进入计算机科学界起到了帮助。然而，由于难以通过应用基准测试反映这种开放式编程模式，在新并行处理器的压力测试中，此类程序及其开发若非完全缺失，往往也未发挥足够作用。这种现实使得传统的、基准测试驱动的硬件设计很容易回避这个领域。

硬件厂商面临着巨大的难题。任何范式转变，例如转向并行所需的范式转变，本身都有风险。厂商在制造此类硬件之前，无法从应用开发人员获得太多反馈，这也无助于减轻风险。通常只有在新硬件可用后才为其开发应用软件，这为厂商们制造了一个先有鸡还是先有蛋的难题。我提到过，过去十年内台式机和笔记本电脑厂商之间的竞争减弱，导致他们没有动力。然而，移动计算和固定计算的不断融合正带来激烈的竞争，随之而来的是为了保持竞争力所需要承担的风险。这对该领域来说是个好消息。

另外的好消息是，硅片面积预算的持续增加和 3D-VLSI 技术的问世，

任何范式转变，例如转向并行所需的范式转变，都存在风险。

加上硬件更大异质性的潜在调整，可能会让厂商增加新的元件，而不需要放弃对当前编程模型的支持。

本文早期版本的一名审稿人用以下有趣的论证质疑本文的基本要点。既然并行计算的唯一目标是性能，每名并行程序员当然都是性能程序员。因此，并行编程的工作量应该用性能（串行）编程的工作量为标准来评判，而串行性能编程本来也比标准的串行编程要费力得多。我的回答是，衡量获得显著（虽然不是最佳）的并行加速比的工作量的标准应该是是获得良好的（虽然并不是最好的）串行性能所需的工作量。这里，良好的串行性能是指计算机学科的学生可以获得的且不需要为新一代机器重新调整代码的性能。换言之：

- 串行编程的性能通常源自编译器带来的优化，这将程序员从为性能而微调代码的大量负担中解脱出来。这样的优化可使程序员更多专注于通过改进算法获得性能，并行编程的情况与之相反。
- 当受到硬件制约时，程序员能做的也就这么多了。当前的计算机体系结构都面向串行代码（线程）多数据流而设计。为了性能，程序员需要提供足够“大”的线程，这在不规则程序中尚不能实现。

尽管如此，该审稿人的意见和我的回答表明，本文还必须证明我不是凭空想象，能在不规则问题上实现较高加速比和可减少工作量的算法的多核系统确实可行。因此，当前版本引用了 Vishkin 描述的 [a]XMT 多核计算机平台，[7] 它为本文提供了有力的证据。例如：

- 我在马里兰大学给研究生讲授并行算法理论课时，要求他们完成五六份复杂的并行编程作业，最终，与其最佳串行版本相比，几乎所有人都取得了显著加速比。
- 近 300 名中学生，其中大多数来自弗吉尼亚州亚历山德里亚市托马斯·杰斐逊科技中学，已经对 XMT 进行了编程并取得了显著的加速比。
- XMT 主页 [b] 也列举了与初中、市中心平民区高中、大学新生和其他本科生在串行计算上取得的同等或几乎同等的成功。
- 对于之前提到的最大流问题，XMT 主页也列举了一篇文献，展示了超过最佳串行算法计数周期 100 多倍的加速比，然而任何商业系统的加速比都不超过 2.5 倍。
- 该主页上还列举了一些展示其他先进并行算法取得类似 XMT 加速比的文献。
- 另外，XMT 还证明了它与串行代码的向后兼容性不存在冲突。

能够有效处理通用编程和应用的商用并行机器尚未面世。

能够有效处理通用编程和应用的商用并行机器尚未面世。研究界的任务是，垂直开发集成计算栈和原型产品，以及使用重要应用和更容易的并行编程对其进行验证，为厂商开拓前进的道路。由于风险级别高，原型产品开发最适合在研究界进行。另一方面，在现今商业系统中利用并行计算符合企业界的直接利益。如果企业界为现今商用系统的应用开发提供资助，当今稀缺的科研经费就能更多用于最需要的原型产品开发。

科学社会学领域的创始人 Ludwik Fleck（以色列裔波兰人）研究发现，研究界的论述并非没有问题。他指出，甚至研究界最基本的共识（例如，什么是事实）也值得质疑。[4] 尤其是要通过研究界外部的反馈达成共识。本文的目的是为通用多核并行计算提供这样的反馈。其理想影响是跳出潜力有限的陈旧路线和临时解决方案，推动该领域走向寻求实现系统性的进步，并取得和这些旧的路线和解决方案曾经一样的重要性和成功。 C

a XMT 代表显式多线程，不应将其与名为 Cray XMT 的 Tera 计算机项目混淆。

b XMT 主页网址为 http://www.umiacs.umd.edu/~vishkin/XMT/。

参考资料

1. Ballard, G. et al.Communication efficient Gaussian elimination with partial pivoting using a shape morphing data layout.In *Proceedings of the 25th ACM Symposium on Parallelism in Algorithms and Architectures* (SPAA), (Montreal, Canada, 2013), 232–240.
2. Borkar, S.Y. et al.Platform 2015:Intel processor and platform evolution for the next decade.White Paper, Intel Corporation, 2005.
3. Edwards, J.A. and Vishkin, U. Parallel algorithms for Burrows-Wheeler compression and decompression. *Theoretical Computer Science*, to appear in 2014; see http://dx.doi.org/10.1016/j.tcs.2013.10.009.
4. Fleck.L. *The Genesis and Development of a Scientific Fact*, (edited by T.J.Trenn and R.K.Merton, foreword by Thomas Kuhn).University of Chicago Press, 1979. English translation of *Entstehung und Entwicklung einer wissenschaftlichen Tatsache.Einführung in die Lehre vom Denkstil und Denkkollektiv* Schwabe und Co., Verlagsbuchhandlung, Basel, 1935.
5. Fuller, S.H. and Millet, L.I., Eds.*The Future of Computing Performance:Game Over or Next Level.* National Research Council of the National Academies, The National Academies Press, 2011.
6. Valiant, L.G.A bridging model for parallel computation. *Commun.ACM 33*, 8 (Aug. 1990), 103–111.
7. Vishkin, U. Using simple abstraction to reinvent computing for parallelism.*Commun.ACM 54*, 1 (Jan. 2011), 75–85.

Uzi Vishkin (vishkin@umd.edu) 是马里兰大学电气与计算机工程系教授，马里兰高级计算机研究所教授。

部分资助来自美国国家科学基金会 CNS-1161857 CCF-0811504 奖。

译文责任编辑：陈文光

DOI:10.1145/2578902

文章编写主导者 acmqueue
queue.acm.org

互联网边缘 – – 放诞不羁。

作者：PAUL VIXIE

速率限制状态

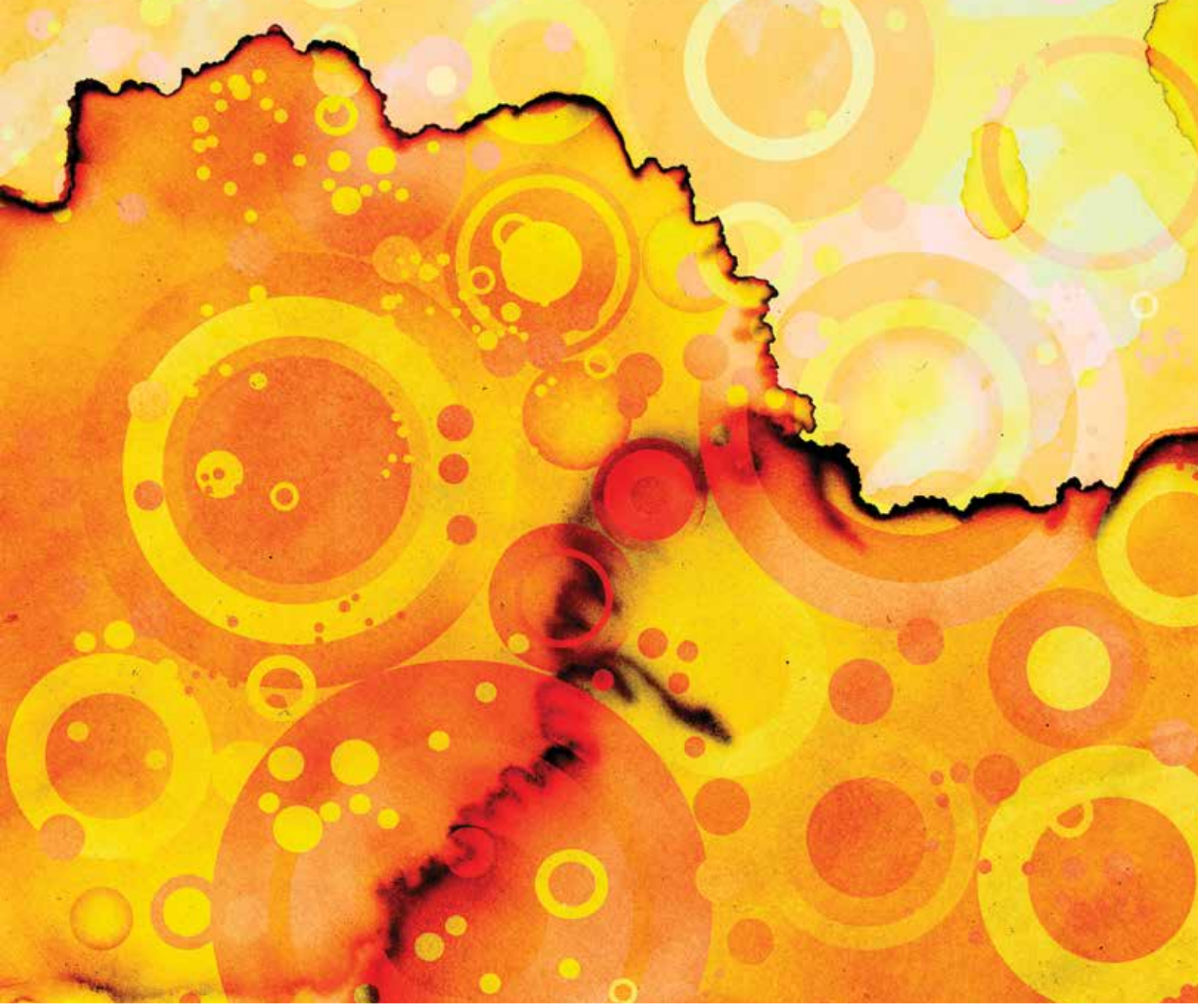
按设计，互联网有着傻瓜式的“*内核*”，智能型的“*边缘*”。正是这种简化内核的设计使互联网的主体能够赶上需求增长的速度，从而得以疯狂发展。将智能性都置于互联网的边缘这一决定有其不利一面，即互联网总体通信质量受其规模的左右。并非所有互联网设备与软件的生产厂商都具有与互联网规模相称的技能以及质保预算。此外，互联网的弹性还意味着在网络通讯方面具有重大错误的设备或程序颇有可能不受自身缺陷的羁绊，而“足够良好地运行”。

现象一：数亿台用户驻地设备 (CPE) 实际上将过多的内存用于缓存数据包。正如 Jim Gettys 和 Dave Taht 近年来所展现的那样，用于数据包的内存并非越多越好。[1] 通信量增加时，家用、公用和商用无线网络的性能全都会以令人惊讶的程度恶化。人们会想当然地认为存在一种“平均分配带宽”的机制，即有 N 路网络通信时，每路通信会大致获得可用带宽的 1/N，但事实上每路通信最终陷入流沙的境地，各获得可用带宽的 $1/(N^2)$。这并非是因为 CPE 设计者能力不足；而是因为互联网是一个涉及大量微妙交互的巨大空间。这些交互依赖于每位设备和软件设计者都做出完全相同的假设，但其中多数假设都尚未约定成文。

现象二：几乎所有智能手机、笔记本电脑、桌面操作系统和应用程序的供应商都源源不断地发布补丁和安全漏洞公告。不怀好意者有时间、技术和动机通过研究边缘设备找出漏洞，并将会找到所需数量的漏洞，以便能往我们的宝贵设备中植入恶意代码。进而在这些设备中复制数据，篡改已安装软件，监控我们，盗用身份 —— 113 年的科幻小说史都未想象出每个人都在线后，我们及我们的生活有多么易受侵犯，从而让我们有所准备。鉴于自由和隐私的侵犯者现已包括国家政府，因此对于全世界来说，边缘设备及其软件的极端脆弱性成为全新的普遍性人权问题。

源地址验证

互联网基本架构中最为可怕的瑕疵在于大多数网关都未能实行简单源地址验证 (SAV)。因为即便没有 SAV，互联网也能运转得很好，因为互联网根植于没有非可信用户和设备的学术界，所以可以确定地说，大部分网关制造商（如无线路由器、DSL 调制解调器以及其他形式的 CPE）都允许多数边缘产品播发任意自称源地址的网络数据包。更糟的是，商业级互联网接入服务



提供者、互联网数据托管中心和各种“云”的运营商也都懒得针对自身客户开启 SAV。运营费用增加可能是原因之一（由为 SAV 会耗费电能，并需要额外的培训和监控），但不默认提供 SAV 的最大原因在于：SAV 只惠及他人客户，而非运营商自己的客户。

从外部无法核查一个网络是否使用了 SAV。任何类型的 SAV 合规测试都必须由位于合规情况存疑网络中的设备来完成。这就意味着，一开始就毫无动力部署 SAV 的网络运营商却是唯一能够告知是否已部署 SAV 的人。这对于 SAV 状况的总体改善并非好兆头，即便有法律或协议的支持。虽然对于保险和核查较为普遍的国家，部署 SAV 或许能成为一项保险和核查要求，但是只要全球大部分地区都没有说服自己重视 SAV 的缘由，那么可以确信地假设，仍有相当大部分的互联网边缘会始终允许数据包层面的源地址伪造，因此我们最好开始学着怎样去适应它 —— 且永远地适应它。

虽然 SAV 的缺位可能产生一些有趣的数据毒化问题，但是目前为止，数据包伪造最为有害之处在于其助长 DDoS（分布式拒绝服务攻击）的方式。[3] 如果任何人都能播发自称来自他人的数据包，那么攻击者可产生一个适量的请求流，并将之指向可公开访问且性能十分强大的互联网服务器，同时造成该请求流来自受害人的假象，这将令受害人疲于应付其从未提交的请求产生的响应。更为糟糕的是，受害人无从追踪到攻击进入网络的位置，因此要么束手无策地等待攻击结束，要么聘请网络安全服务商来减轻攻击，以便受害人的其它服务仍能在攻击期间如常提供。[2]

域名系统响应速率限制

在数年前的一波攻击中，性能十分强大的公共 DNS（域名系统）服务器被用于反射和放大一些有效的 DDoS 攻击。对此，互联网研究者

图片来自 SHUTTERSTOCK.COM

Paul Vixie 和 Vernon Schryver 开发出一套名为“响应速率限制 (DNS RRL)”的系统，让被利用进行这些反射 / 放大攻击的 DNS 服务器运营商能够有意地丢弃从统计角度看来与攻击相关的整个输入请求流集合。[4] DNS RRL 并非一个完美的解决方案，因为其可能会在攻击期间令占少数的正常（非攻击）事务略有延迟。然而，鉴于所有现代 DNS 服务器，甚至一些入侵防御系统 / 入侵侦测系统 (IPS/IDS) 产品现在都采用某种形式的 DNS RRL，以及许多顶级域 (TLD) DNS 服务器正在运行 DNS RRL，因此 DNS RRL 的这些取舍显然是有益的。高性能互联网服务器的运营商必须都学会和遵照 Stan Lee 设定的规则（如蜘蛛侠所说）：“实力愈强，责任越大”。

DNS RRL 曾是一个因域而异的解决方案，取决于对 DNS 本身的深入了解。例如，DNS RRL 限制*响应*速率的原因，在于仅问题出现这一情况不足以让速率限制系统判断某项请求是否属于某次攻击。然而，鉴于还能做前瞻性响应，因此可较确信地侦测到存疑的伪来源，从而减少 DNS 服务器被用作 DDoS 反射放大器的情况，使之仍然能为同时发生的非攻击通信提供“足够好的服务” —— 即便这些非攻击通信与攻击十分相似。

信息战的经济学与任何其它类型战争的经济学无异 —— 既寻求低于攻方攻击成本的防御成本，也寻求低于守方防御成本的攻击成本。在过去，DNS RRL 并不需要完美无缺，仅需要能打破平衡：使 DNS 服务器对攻击者的吸引力低于其可用的替代手段。DDS RRL 的一个重要设计原则是，将 DNS 服务器变成 DDoS *衰减器* —— 即不仅仅消除放大效果，而且实际的攻击量少于攻击者可通过直接发送数据包取得的攻击量。同样重要的是，这种衰减不仅能以字节 / 秒为单位计，而且能以数据包 / 秒为单位计。在一个满是复杂状态防火墙的世界中，这点十分重要。在这样的环境中，瓶颈往往是数据包的数量，而非字节的数量，而且按防火墙性能来计，处理小数据包的成本与处理较大数据包的成本相当。

不怀好意者有时间、技术和动机通过研究边缘设备找出漏洞，并将会找到所需数量的漏洞，以便能往我们的宝贵设备中植入恶意代码。

DNS RRL 设计的另一项重要标准是运行成本低至不值得一算的程度。DNS RRL 使用的 CPU 运算量、内存带宽、内存空间仅在 DNS 服务器的总体负荷中占极小比例，以至于攻击者无法以某种方式令 DNS 服务的 RRL 性能“超载”，从而打消 DNS RRL 对服务器运营商的吸引力。再次强调，战争是一种应用经济学，DNS RRL 的设计特定地将守方的成本限制成攻方一*小部分*成本的一小部分。鉴于 DNS 通过不分状态实现了卓越的性能和伸缩性，因此 DNS RRL 应尽可能地减少了因阻止反射放大攻击而需要向 DNS 添加的状态数量，进而不降低 DNS 的性能。

现状

如要在各种网络协议的背景下不分状态，这就意味着响应设备不需要在两次请求之间记住与请求设备相关的任何信息。每个请求都完全独立。对于 DNS 来说，这意味着一个请求进来，一个响应出去，就像一次返还于请求设备和响应设备之间的单次双程旅行。并不禁止响应者有选择地添加状态 —— 例如，DNS RRL 添加某种适当的状态，以帮助区分攻击数据包与非攻击数据包。请求者也可选择对每台候选服务器嵌入诸如 RTT（往返时间）等状态，这样可以将今后的事务导向响应最快的服务器。然而，在 DNS 中所有此类状态都非强制性的，即便任何一端都不记录任何状态，该协议本身也能运转如常。

DNS 是用户数据报协议 (UDP) 的一个实例；类似的协议还有一些。比如说，网络时间协议 (NTP) 采用

UDP，并且每次响应的大小都等于或大于相应请求的大小。一个真正的 NTP 客户端存有某些状态，用以追踪互联网时间。然而，攻击者无需向 NTP 响应设备出示表明存在此类状态的任何证据，也能引发响应。由于 CPE 网关或者其他边缘设备常常嵌有 NTP，因此有数千万响应设备可被 DDoS 攻击者用作反射器或者具有放大效果的反射器。

另一方面，传输控制协议 (TCP) 是区分状态的。在当前设计中，请求发起设备和响应设备都必须记住对方的一些信息；否则，通信无法实现。区分状态有利有弊。如要发出一个请求并收到一个响应，需要数次往返才能在两边确立足够的连接状态，然后还要 1.5 次往返才能关闭连接且释放两端的全部相关状态，因此这会造成沉重的负担。TCP 有一个尝试在两个端点之间建立共享状态的初始化期。期间响应设备可向单个 SYN（同步）消息的声称发起设备发送数个 SYN-ACK（同步确认）消息。这意味着即便 SYN-ACK（同步确认）消息并不接连（背靠背地）发送，TCP 本身也可用作字节和数据包的放大器。鉴于有数亿个可供利用的 TCP 响应设备，因此无论受害人拥有多么丰富的资源或者多么完善的防御，DDoS 攻击者可轻松地找齐发起任意攻击所需的 TCP 反射放大设备。

互联网控制消息协议 (ICMP) 是不分状态的，因为网关和响应设备根据网络情况和发起设备的行为，以异步响应方式向发起设备传回消息。常用的“ping”和“traceroute”命令依赖广泛可用的 ICMP；因此，鲜有防火墙阻止 ICMP 的情况。每个互联网网关和主机都以某种形式支持 ICMP，因此基于 ICMP 的 DDoS 反射攻击者可以找到他们期望数量的 ICMP 反射器。

贯穿以上结论的主旨是在缺少 SAV 的情况下，不分状态是很糟糕的。虽然许多其他基于 UDP 的协议（包括服务器消息阻止 (SMB) 和 NFS）在运用得当时是区分状态的，但是与 TCP 一样，会在初始连接建立期不分状态，因而视 DDoS 攻击者的技术水平，会被用作 DDoS 反射器或 DDoS 放大反射器。虽然所有这些问题的根本原因是长期普遍性地缺乏 SAV，但其近因是那些不分状态的协议。显然，要在一个没有 SAV 的世界中生存，互联网、每种协议、每个系统都将需要更多地运用状态。这种状态区分并不会进入永远傻瓜式的互联网内核。相反，为能应对无 SAV 的情况而必须强加给互联网系统的状态只能在互联网边缘添加。

结论

本文提及的各类有助于反射的协议都将需要学会速率限制。这包括初始的 TCP 三次握手，ICMP 和每个基于 UDP 的协议。虽然在少数情况下利用防火墙来限制某个设备对 DDoS 反射和 / 或放大的参与，但是大多数防火墙都要么本身不区分状态，要么区分状态能力弱到会被逐个击破。更为常见的情况将与 DNS RRL 一样，即需要对协议本身有较深的了解，才能正确地设计出适用于该协议的速率限制解决方案。工程经济学要求，为速率限制而添加的任何新状态所产生的 CPU、内存带宽以及内存空间成本，相对于攻击者的努力都应该是微不足道的。衰减也必须是第一位的目标 —— 我们必须让直接给受害人发送数据包比通过 DDoS 衰减器反弹对攻击者更具吸引力。

这种努力将会历经数年，耗资巨大。虽然这会比 SAV 贵得多，但是 SAV 因人们动力不对称而完全不具可行性。通适协议感知速率限制（采用 DNS RRL 方式，但旨在用于互联网上的所有其他目前不分状态的交互）有以下激励模式的绝佳优势：需要从事某工作的人确有动机来完成这项工作。这一努力是互联网这种“傻瓜内核、智能边缘”模型以及伯斯塔尔法则（Postel’s Law）（“稳健行事，广纳众言”）所必然产生的成本。

随着互联网群体规模的增长，反射式和放大式 DDoS 攻击呈稳定增长趋势。越多的受害人以新方式依赖互联网，DDoS 攻击者的动力就越足；而且越多的创新者将越多的智能设备添加到互联网边缘，发起 DDoS 攻击的成本就越低。既没有办法能令 SAV 普及到有重要影响的程度，也没有办法在 SAV 以某种方式神奇地成为强制要求后，集中地衡量或核查合规情况。

DDoS 将继续增加，直至互联网变得十分拥挤不堪，以至于新 DDoS 攻击发起者所得到的好处仅相当于噪声，也就是说直到包括攻击者在内的所有人都淹没在网络噪声之中。另一个选择是，为互联网上每一个目前不分状态的协议、服务、设备添加速率限制状态。 C

queue.acm.org 上的 相关文章

DNS Complexity
Paul Vixie
http://queue.acm.org/detail.cfm?id=1242499

Broadcast Messaging:
Messaging to the Masses
Frank Jania
http://queue.acm.org/detail.cfm?id=966719

Lessons from the Letter
George V. Neville-Neil
http://queue.acm.org/detail.cfm?id=1837255

参考资料

1. Bufferbloat; http://www.bufferbloat.net/.
2. Defense.net; http://defense.net/.
3. Vixie, P. Securing the edge, 2002; http://archive.icann.org/en/committees/security/sac004.txt.
4. Vixie, P. and Schryver, V. Response rate limiting in the Domain Name System, 2012; http://www.redbarn.org/dns/ratelimits.

Paul Vixie 是 Farsight Security 公司首席执行官，互联网系统联盟（ISC）的创建者，曾任 ISC 总裁、董事长；曾任 MAPS、PAIX 和 MIBH 的总裁以及 Abovenet/MFN 的首席技术官（CTO）。Vixie 是 ICANN RSSAC（根服务器系统咨询委员会）和 SSAC（安全和稳定性咨询委员）的创会会员。

译文责任编辑：陈贵海

DOI:10.1145/2500876

利用源于开创性的 MITx（现为 edX）课程 -《6.002x: 电路与电子学》的学生参与数据：解读 MOOC 中的学生行为。

DANIEL T. SEATON 、 YOAV BERGNER、ISAAC CHUANG、PIOTR MITROS 和 DAVID E. PRITCHARD

大规模在线公开课程中的人与事

大规模在线公开课程（MOOC）收集了学生学习行为方面的珍贵数据；这些数据涵盖了学生在独立自足的学习环境中所有互动的完整记录，还具有大样本的优势。我们在文章中总结了 108,000 位《6.002x - 电路与电子学》课程参与者的行为。6.002x 是 MITx（现为 edX）在 2012 年春季学期开设的第一门课程。我们按测评活动的参与程度把参与者分为不同的群体，其中包含了从浏览者（占参与者总数约 76%，但只占课程总学习时间的 8%）到证书获得者（占参与人数的 7%，但占总时间的 60%）之间的各种群体。我们考查了证书获得者如何在不同的课程要素之间分配时间，也审视了各种要素中被证书获得者访问的部分。我们还分析了课程要素之间的过渡，阐释了解答作业习题与考试问题时学生行为的不同之处。本次研究为继续研究课程要素及其过渡对 MOOC 学习的影响奠定了基础。

虽然出现的时间并不短，[8] 但是到了 2011 年末后，免费在线课程方才达到了前所未有的规模。三大组织 - Coursera， edX 和 Udacity- 均发布了 MOOC[13]，其中每门课程的注册人数达到了 100,000 名。自从 MOOC 发布后，三大初创组织的各项数据均有增长，课程总数已经过百，注册总人数也到了三百万。据此，《*纽约时报*》把 2012 年誉为“MOOC 年”。[16] 虽然业界对这些初创组织重塑高等教育的方式怀有诸多疑问，[6,12,20] 但迄今为止，却鲜见分析和描述其中学生行为或学习情况的论文发表。

MOOC 中存在可以利用的海量数据，我们的主要目的则旨在阐明如何利用这些数据创造独特的研究机会和手段，分析和解读在独立自足的学习环境下，学生在整个课程中的详细行为。据此，我们研究了约 100GB 的日志数据。这些带有时间戳的数据描述了 2012 年春季开设的开创性 MITx 课程《6.002x 电路与电子学》中的学生互动情况。与以前的在线学习研究相比，此次的数据至少大了两个数量级。[10,21]。此外，我们还开发并阐述了多种研究学生与课程资源互动的方法。虽

» 重要见解

- **从 MOOC 中收集的数据可让我们深入了解学生行为。数据范围涵盖了从每周阅读电子教科书的习惯到解答习题时各种行为。在这些行为中，学习资源的使用会因场景不同而有所变化。**
- **在 6.002x 课程中，76% 的参与者是浏览者，他们在课程上花的时间加起来只占总时间的 8%。与此相反，努力获取证书的参与者占人数的 7%，每人平均花了 100 个小时，加起来占总时间的 60%。**
- **学生每周的大多数时间花在了教学视频和作业方面的互动上，讨论论坛和在线实验室次之；然而，视频和课堂提问方面的互动呈现明显的双峰，其中半数证书获得者的访问量不足这些资源数量的一半。**

插图由 ANTHONY FREDA 提供

然我们没有分析人口因素，不过我们却从学生尝试的测评项目数和学生花费在课程上的总时间两方面对学生进行了区分。在使用这些度量指标研究了所有的注册学生后，我们把关注点转到了获得结业证书的学生上，深入研究他们的时间分配和资源使用情况。对于证书获得者，我们从用户的时间分配及其访问的所有细节部分两个视角探讨了各种课程要素（如教学视频、作业和讨论论坛）的使用情况。我们还研究了解答习题时的资源使用情况。研究表明，与参加考试相比，解答作业习题时，学生的访问模式及其在不同课程要素之间的时间分配均存在显著的差异。

退学率与开始数周中花费的时间更少之间具有相关关系。依据这种关系是否可以推断，激励学生投入更多的时间后，保留率会有所升高。

6.002x 过程、数据分析

为在线使用而进行稍许修改后，为期 14 周单元的 6.002x 课程无论是从格式还是教学进度上都大致模仿了传统的校内课程。课程计划（见左侧导航栏图 1）包含课堂计划，其中包括教学视频（带注释的 PowerPoint 胶片和实际的 MIT 讲座）。在教学视频中，内置了课堂提问、教程视频（替代背诵）、作业（三至四个分为多个部分的习题）以及试验任务（互动的电路工具箱）。总成绩根据作业（15%）、试验（15%）、期中考试（30%）和期末考试（40%）的成绩综合评定。补充材料（见图 1 的顶部导航条）包括课程教科书（可自由跳转的页面图像）、员工和学生均可编辑的维基以及主持的学生讨论。如果您想了解课程结构及可用资源的详细信息，请访问位于 https://6002x.mitx.mit.edu/ 的课程档案。

分析跟踪日志。分析跟踪日志是一种了解混合课程和在线课程中学生行为的惯用手段。[5,14] 在 6.002x 的跟踪日志中，每种互动（点击）包含了各种相关信息，其中包括用户名、资源 ID、互动详情和时间戳。互动详情是场景相关的（例如提交的作业习题回答的正确性，讨论帖的正文文本，翻阅的书籍页码）。edX 软件是通过云进行发布的；也就是说，互动数据保存在多个服务器上。总的来说，在最初的 2012 年春季学期，共有 38,000 个日志文件，记录了约 2 亿 3 千万次互动。

我们对日志进行了预处理，将其按每个参与者分割为独立的时间序列，然后统计参与者级的描述性资源使用数据，其中包括访问的资源数量（不计重复），每种资源类型的总访问频率以及每种资源的总使用时间。我们还分析了习题答案的提交情况，并生成了一个响应矩阵描绘回答的正确性和数量。在可行的情况下，我们会利用支撑 6.002x 课件的 MySQL 数据库核对事件日志的访问数据。所有的日志分析工作均由 Python 和 R 中的标准模块来完成。

资源的使用时间估算估算每位课程参与者的时间时，我们利用了学生开始与资源的进行互动的时间以及学生转到其他地方时的时间，求出其中的持续时间。我们从每位参与者的时间序列中求出持续时间，然后累计这些持续时间，获取每种单独的课程要素（包括作业、教科书和讨论论坛）的时间。我们发现有证据表明，如果持续时间小于三秒，则表示学生已经转向了所需的资源界面；因此，我们没有把这些时间片段当成活动。另外，我们也没有把超过一小时的持续时间累计在内，因为我们假设用户在此期间已经离开了电脑。使用其他的高截止值（20 分钟至 1 个小时）后，总体的时间会有 10%-20% 的变化，但却不会显著改变课程要素之间的时间分配关系，也不会明显影响不同参与者的总使用时间。

还有一点较为重要，即累计时间关联了当时展现的资源；例如，如果学生做作业时参考了教科书，则累计该时间段时会包含阅读教科书的时间。不过，对于我们而言，我们只在与作业的直接互动中记录作业资源。除了这种方法之外，显然还有其他的可选方法（比如把打开和解答习题之间的所有时间计为解答习题的时间 [21]）。如果用户打开了多个浏览器窗口或者页签，那么我们的时间累计算法就会出现部分失真；为了处理这一问题，edX

的开发人员正在考虑在将来利用其他的手段。

结果

2012 年初，透着新奇，拥着公众的关注，MOOC 吸引了大量的注册者，不过那些注册者只是好奇，却不够认真。现在，我们仍把是否参与测评作为衡量认真度的指标。2012 年春季的 6.002x 课程吸引了 154,000 名注册者，其中有 46,000 名注册者从未访问课程；对于所有其他的参与者，其使用时间的中位数也仅有一小时（见图 2a）。我们原来希望总使用时间呈现双峰分布，其中有一个“浏览者”的大峰值，使用时间只有一小时左右；另一个峰值为证书获得者，使用时间则多于 50 个小时。然而，实际上在这两个峰值之间并没有出现最低值，只有一个明显的肩（见图 2a）。尝试者的使用时间在整个课程使用时间中居中；按被尝试的作业和考试的测评项目数，我们把尝试者分成了不同的群体（用不同的颜色区分）：浏览者（灰色）尝试的项目 < 5% 的

图 1. 6.002x 中常见的学生视图截屏

从下面的界面中可以访问所有的课程要素。左边的侧边栏定义了课程计划；在每周的单元中，包括了课堂计划（视频和提问）、作业、试验和教程。顶端的导航栏提供了补充材料的访问链接，包括数字化教科书、讨论论坛和维基。主框架内展现了第一课的课堂计划；顶端下方的米色方形区域展现了教学视频和提问。

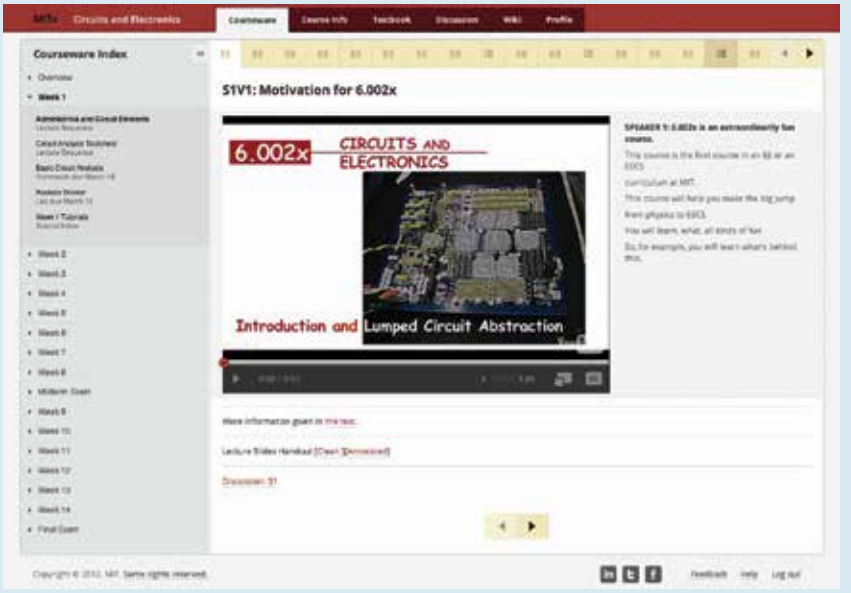


图 2 群体、总时间和退学率

（a）参与者花费在 6.002x 课程上的时间分布（时间轴已经经过了对数转换）；我们把没有获得证书的参与者按其尝试的测评活动的百分比分为多个群体（见表 1）；（b）测出的每个群体花费的总时间的百分比；以及（c）学生每周投入的平均时间。第 8 周和第 14 周左右的阴影区域代表期中考试和期末考试的时间段。

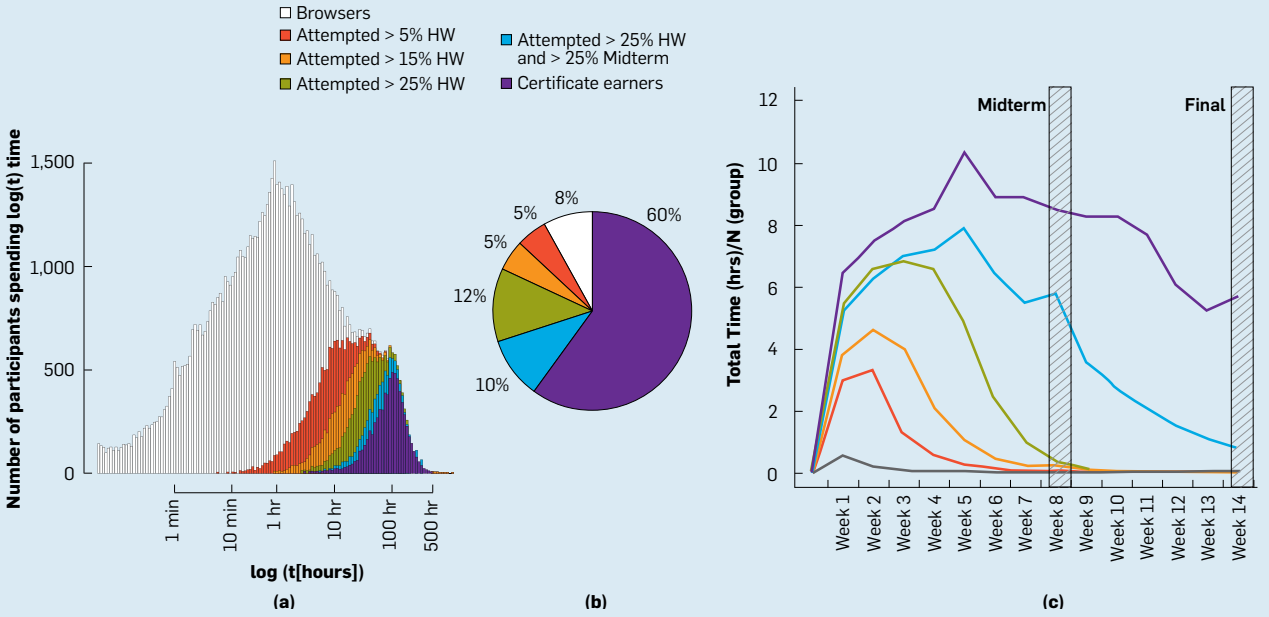
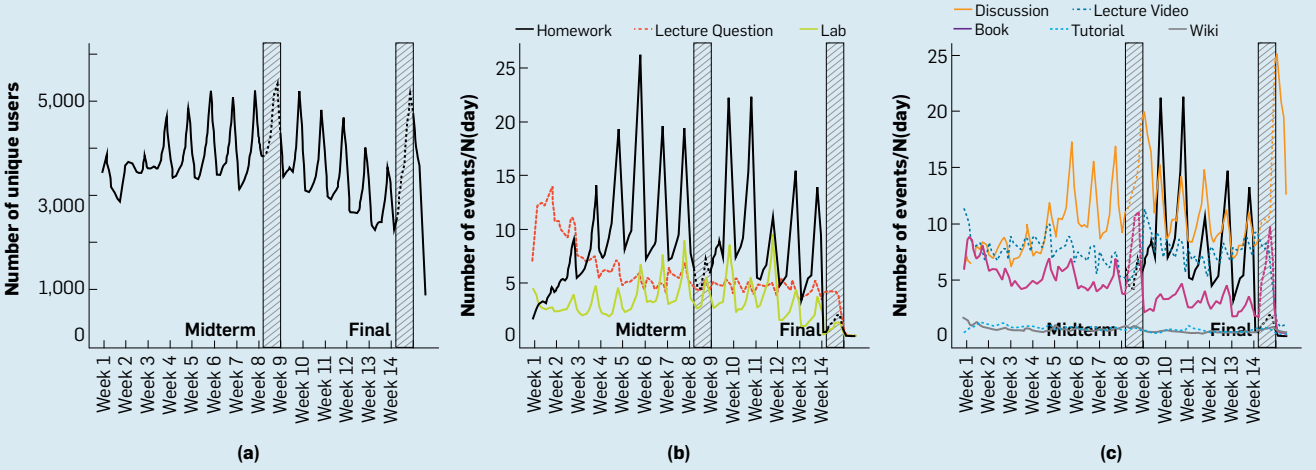


图 3 访问频率

从左到右来看，依次为每日活跃的证书获得者 N（不计重复），其每日平均访问的基于测评的课程要素和基于学习的课程要素的数量。图（a）突出了证书获得者的周期性和趋势。图（b）用于说明测评情况，包括作业、试验和课堂提问，展示了当日活动用户的访问量。图（c）中，基于学习的要素包括教学视频、教科书、讨论、教程和维基。从图中我们可以看出，到了学期后期，讨论论坛用得更多，出现了相当强的周期性。这点与图（a）中可拿分的互动类似。然而，其他要素却缺乏周期性，访问频率也千差万别。第 8 周和第 14 周左右的阴影区域代表期中考试和期末考试的时间段。



作业；群体 1（红色）5%-15% 的家庭作业；群体 2（桔色）15%-25% 的作业；群体 3（绿色）> 25% 的作业；以及群体 4（青色）>25% 的作业和 25% 的期中考试内容。证书获得者（紫色）尝试了大多数所提供的作业、期中考试和期末考试。各群体的课程总使用时间的中位数分别为 0.4 小时、6.4 小时、13.1 小时、30.0 小时、53.0 小时和 95.1 小时。除了上述群体外，150 多位证书获得者花费了不足 10 个小时学习课程。他们可能代表了技术娴熟且想获得证书的群体。与此类似，250 多位考生花费了不足 10 个小时学习课程并完成了多于 25% 的期中和期末考试内容，但却没有获得证书。

图 2c 描述了各群体的参与者每周花费的平均时长，单位为小时。对于尝试测评项目偏少的群体，由于大多数参与者实际上已经退学，他们的曲线不仅降低得更早，而且与证书获得者相比，他们在开始的几周投入的时间也更少。退学率与开始数周中花费的时间更少之间具有相关关系。依据这种关系，值得探讨是否在激励学生投入更多的时间后，保留率会有所升高。

在本文中的其他部分中，我们会聚焦于证书获得者，因为他们占了资源使用的大头；同时，我们还会研究整个学期范围内的时间和资源使用情况。

访问频率。图 3a 描绘了证书获得者中每日的活动用户数量，其中大的峰值出现在星期天，因为星期天是提交用于评分的作业和试验的截止日期；不过，对于课堂提问，情况并非如此。在期中考试和期末考试之间的数周（阴影区域）中，存在下降的趋势。虽然期末考试前两周并未布置任何作业或试验，但是峰值仍然保持稳定。在图 3b 和图 3c 中，对于基于测评的课程要素和基于学习的课程要素，我们分别绘制了每天每位活动学生在不同事件中的活动曲线（点击数会因时间截止值不同而有所变化）。作业组和讨论论坛占每位学生的活动中占比最高。而且在该学期中，讨论活动逐步增加。随着作业活动逐渐增加，课堂提问事件的衰减开始变早。考试时教科书的使用达到峰值；期中考试后，教科书的活动明显减少，这点与传统课程的惯常情况相同。[18]

任务时间。由于时间是学生的主要成本函数，所以研究学生如何在可用的课程要素之间分配时间相当重要。[15,19] 图 4 表明，学生的大部分时间花费在教学视频上；花费的时间为每周三至四个小时，接近于每周安排的视频总时长，所以重新观看和复习视频的学生在时间上必定补偿了那些加速回放或者不观看视频的学生。

在前七周中，最为显著的变化是：时间分配的重心从课堂提问明显转向了作业，如图 4 所示。考虑到以成绩为目标的导向（见图 5），我们应该注意到作业是课程评分的一部分，但课堂提问却不是。不仅如此，即便以掌握技能为目的，学生或许也会觉得，如果没有完成作业，那就无法证明他们已经理解了课程内容。学生在讨论论坛中花费的时间较为突出，这点尤其值得注意，因为论坛既不是课程计划的一部分，也不贡献任何分数。学生在

讨论论坛中花费时间大概是因为论坛所提供在教育或社交上的实用价值，或者两者兼有。期中考试时，花在教科书上的时间呈现出小高峰，访问量也出现了较高的高峰，如图 3 所示。而且，期中考试后，教科书的使用下降。在线资源与传统校内课程混合使用时，教科书的使用也普遍出现这种情况。[18]。在后续研究中有人比较了混合型课程与在线课程中的教科书使用情况，这些研究也比较切合我们的情况。[3,17]

课程要素的使用百分比。除了学生的时间分配之外，各种课程要素的部分使用仍然是一个重要的度量指标，它能帮助讲师确定改进课程的方法，辅助研究者研究课程结构对学生活动和学习的影响。在部分使用方面，我们绘制了证书获得者的百分比图。图中的证书获得者访问了一定百分比的课程要素资源（见图 5）。在作业和试验（各占总成绩的 15%）中，部分使用的情况比较普遍。这些曲线的拐点在 80% 附近。倘若课程策略没有规定需要去掉评分最低的两次任务，比例可能会更高。虽然 6.002x 课程的教学大纲中布置了教科书，但是教科书和教程的使用比例仍然相对不高。在大型物理学入门课程 [16] 的补充性（没有明确地包含在课程计划内）电子文本中，我们也观察到了类似的分布情况。教程视频的使用不多，课程作者会觉得有点失望，怀疑这部分是因为教程在课程计划中的位置不佳。因为在课程计划中，教程被放在作业和试验后面（教程原本用于提供指导）。（由于没有为维基和讨论论坛定义资源数量，所以本处未包含其在内。）

为了更好地解读中间代表教学视频和课堂习题的曲线，想到曲线的负斜率指访问课程要素该部分的学生密度（见图 5b 和图 5c）后，我们便可以利用这一点。有趣的是，教学视频的分布呈现了明显的双峰。76% 的学生访问了超过 20%

图 4. 任务时间

证书获得者平均每周花费在各课程要素上的时间（单位：小时）；阴影部分指该周时举行了期中考试或期末考试。

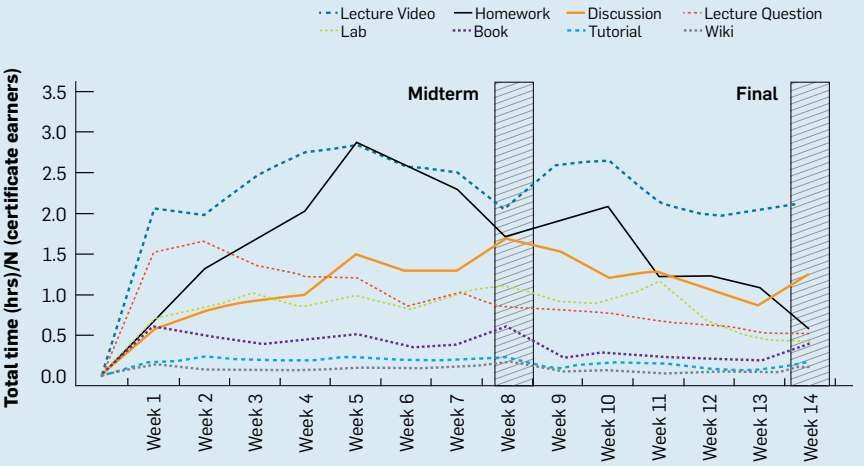
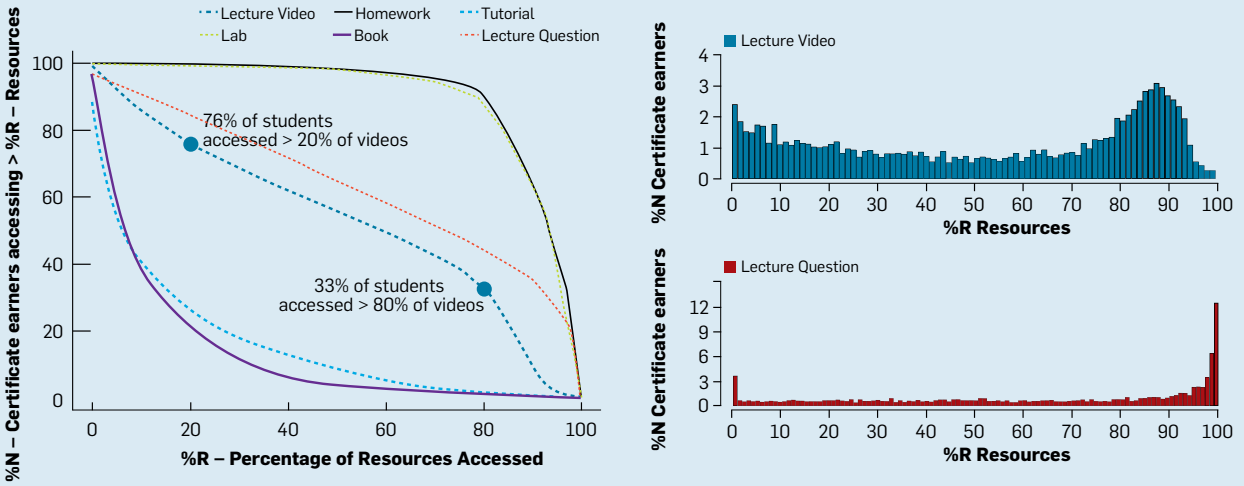


图 5 资源的部分使用

（a）证书获得者（访问的特定课程资源类型的数量高于 %R）的比例用户密度为使用曲线的负斜率。图中标出了用于说明教学视频双峰分布的两个点：76% 的学生访问了 >20% 的教学视频，以及 33% 的学生访问了 >80% 的教学视频。（b）访问的教学视频的双峰分布（百分比）。以及（c）访问的课堂提问的分布



的视频（或者说 24% 的学生访问的资源数量低于 20%），33% 的学生访问了超过 80% 的视频。这种双峰的形态说明我们应该继续研究学习偏好；例如，有些学生在学习时是否只用了其他的资源？或者说，他们是否在注册课程前已经掌握了课程内容？课堂习题的使用在 0% 至 80% 的区间内平坦分布，然后暴增，说明很多学生访问了几乎所有的教学习题。在上半个学期，花费在课堂提问上的时间稳步下降（见图 2）。除了上述实际情况之外，此分布还意味着学生不仅减少了回答课堂提问的时间，有些甚至完全不回答课堂习题。

解答习题时使用的资源。学生对资源的顺序使用存在多种模式。在这些模式中，可能会包含认知乃至情感状态方面的线索。[2] 因此，我们可以把时间序列数据转换为资源间的过渡矩阵，并通过这些矩阵来探索测评的使用和学习资源之间的相互作用。过渡矩阵包含了所有单独的资源与资源间的过渡，我们把这些过渡累计起来作为主要课程要素之间的过渡。6.002x 的学习环境相当完备，也就是说学生在参阅教科书、复习以前的作业，或者搜索讨论论坛时，都不需要离开这个环境。因此，我们也拥有了难得的机会，可以观察学生在解答习题时访问的所有课程要素及其之间的过渡。以前研究在线解答习题时，正好缺乏此类信息。[21]

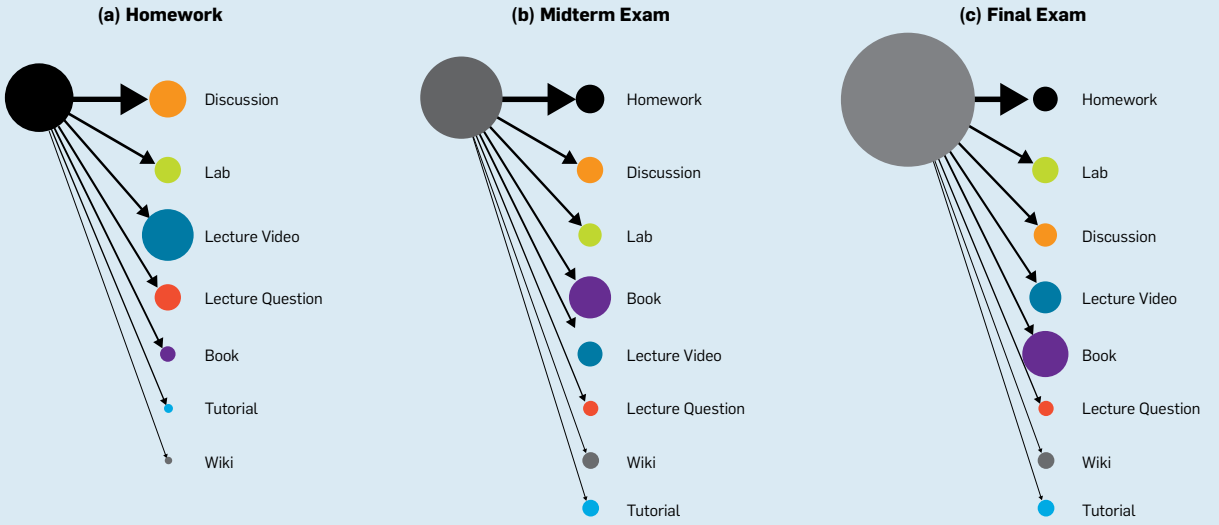
图 6 突出了学生从习题（解答习题时）到其他课程要素的过渡。图中，作业、期中考试及期末考试作为目标各异的测评类型。图 6 还说明，虽然教学视频占了多数时间，但在解答作业中的习题时，讨论论坛却是最常见的目的地。在考试中（期中和期末考试的情况类似），以前完成的作业为主要的着眼点，但大多数的时间却用在了教科书上。与解答作业中的习题相比，解答考试问题的学生行为截然不同。注意，因为作业都放在一起，我们无法把“参阅之前的任务”从学生完成作业的活动中分离出来。

结语

本文在课程分析方面做出的主要贡献在于说明了可用多种不同的方法定性分析 MOOC 数据以解决下列重要问题：退学率 / 保持率、学生时间在资源之间的分配方式、资源的部分使用以及解答系习题时的资源使用情况。本文所展示较为显著的结果为：虽然只有 25% 的参与者尝试了超过 5% 的作业，但是他们花费的时间却占了花费的课程总时间的 92%；实际上，6% 的参与者最后获得了证书，但他们投入的时间却占了总时间的 60%。与证书获得者相比，放弃课程学习的参与者投入的精力也较少。那些在前两周投入精力最少的参与者放弃课程的时间往往会更早。对于大多数证书获得者，他们把大部分时间花费在教学视频上，但也有约 25% 的证书获得者观看的教学视频低于 20%。这种情况说明我们需要进行后续调查，研究资源使用与学习之间的相关性。最后，我们强调了，虽然在导航序列中未要求，也未包含讨论论坛，但是讨论论坛的流行程度仍然相当明显。倘若这种社交性的学习要素在 6.002x 的成功中扮演了重要角色，那么完全异步的其他方式可能会不够吸引人。至少对于电路与电子学之类的复杂课程话题来说，情况如此。

在校内课程研究中，业界探讨了在（大学）入门课程中，课程结构对资源使用 [18] 和最终成绩 [4,11,19] 的影响方式。在本次研究中，某些结果与那些校内课程研究中发现的结果相同。本研究和 MOOC 的后续研究应该能从总体上为校内教育提供启发。另一方面，MOOC 或许可以利用现有校内教育的研究成果（比如频繁的考试不仅可以提高资

图 6 解答（a）作业，（b）期中考试和（c）期末考试中的习题时，转向其他要素的过渡。箭头的宽度与过渡的总体数量成正比，而课程要素则按箭头宽度从上到下依次排列；节点的大小由花费在该要素上的总时间决定。



源的使用，还能尽量提高学生的学习成果[11]）。

最后，我们强调，MOCC 提供了独特的视角来观察数量庞大、背景各异的学生群体的学习情况，可以让大家深入了解课程的方方面面，然后据此进行研究。与大多数以前对校内教育环境的研究相比，我们拥有了大致囊括所有学生行为和整门课程中相关学习情况的日志，这些日志都带有时间戳。所有这些数据都带有可靠的统计信息，并可用于研究特定的学生群体（比如根据付出的精力、学习习惯和人口背景信息划分的群体[9]）。把任务时间的观察数据与学习的度量数据结合在一起后，我们开拓了一条度量学习价值（在给定课程要素上花费单位时间后掌握的知识量）的途径，或许还拓展了以前的在线学习研究。[7,15]。本研究反过来又会推动基于研究进展、试验和学习成果的度量等因素进行螺旋式的革新，进而改进教育内容及其传递方式。因为很多 MOOC 课程在资源类型、格式以及课程计划方面均大致模仿了传统的校内课程，我们希望我们的成果也能为深入了解和改进传统校内课程中的学习提供一些借鉴。

鸣谢

本文得到了美国国家科学基金会的支持，项目资助的编号为 DUE（本科教育部）-1044294，但是本文观点与美国国家科学基金会无关；此外，Google 奖教金也提供了资助。我们在此感谢 MITx 给了我们相关数据的访问；特此感谢 J·德波尔以及 MIT 的教学和学习实验室（Teaching and Learning Laboratory）和高效学习、测评和辅导（Learning, Assessing and Tutoring Effectively ）研究所的各位同仁，他们的建议和观点让我们受益匪浅。

在考试时，教科书的使用达到峰值；期中考试后，教科书的活动明显减少，在传统课程中也普遍存在这种情况。

参考资料

1. Ames, C. and Archer, J. Achievement goals in the classroom:Students' learning strategies and motivation processes.*Journal of Educational Psychology 80*, 3 (1988), 260.
2. Baker, R.S., D' Mello, S.K., Rodrigo, M.M.T., and Graesser, A.C. Better to be frustrated than bored:The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies 68*, 4 (2010).
3. Cummings, K., French, T., and Cooney, P.J.Student textbook use in introductory physics.In *Proceedings of the Physics Education Research Conference*, 2002.
4. Freeman, S., Haak, D., and Wenderoth, M.P. Increased course structure improves performance in introductory biology.*CBE-Life Sciences Education 10*, 2 (2011).
5. Guzdial, M. *Deriving Software Usage Patterns from Log Files.* Technical Report GIT-GVU-93-41, 1993.
6. Hyman, P. In the year of disruptive education. *Commun.ACM, 55*, 12 (Dec. 2012).
7. Jiang, L., Elen, J., and Clarebout, G. The relationships between learner variables, tool-usage behavior, and performance.*Computers in Human Behavior 25*, 2 (2009).
8. Johnstone, S.M.Open educational resources serve the world.*Educause Quarterly 28*, 3 (2005).
9. Kolowich, S. Who takes MOOCs?*Inside Higher Education 5* (2012).
10. Kortemeyer, G. Gender differences in the use of an online homework system in an introductory physics course.*Physical Review Special Topics:Physics Education Research 5*, 1 (2009).
11. Laverty, J.T., Bauer, W., Kortemeyer, G., and Westfall, G. Want to reduce guessing and cheating while making students happier?Give more exams!*Physics Teacher 50*, 9 (2012).
12. Martin, F.G.Will massive open online courses change how we teach?*Commun.ACM 55*, 8 (2012).
13. McAuley, A., Stewart, B., Siemens, G., and Cormier, D. The MOOC model for digital practice.Social Sciences and Humanities Research Council, *Knowledge Synthesis Grant on the Digital Economy*, 2010.
14. Minaei-Bidgoli, B., Kortemeyer, G., and Punch, W.F.Enhancing online learning performance:An application of data mining methods. *Immunohematology 62*, 150 (2004).
15. Morote, E.S. and Pritchard, D.E.What course elements correlate with improvement on tests in introductory Newtonian mechanics?*American Journal of Physics 77* (2009).
16. Pappano, L. The year of the MOOC. *The New York Times* (Nov. 2, 2012).
17. Podolefsky, N. and Finkelstein, N. The perceived value of college physics textbooks:Students and instructors may not see eye to eye.*The Physics Teacher 44* (2006).
18. Seaton, D.T., Bergner, Y., Kortemeyer, G., Rayyan, S., Chuang, I., and Pritchard, D.E.The impact of course structure on etext use in large-lecture introductory-physics courses.In *Proceedings of the Physics Education Research Conference*, 2013.
19. Stewart, J., Stewart, G., and Taylor, J. Using time-on-task measurements to understand student performance in a physics class:A four-year study. *Physical Review Special Topics-Physics Education Research 8*, 1 (2012).
20. Vardi, M.Y.Will MOOCs destroy academia?*Commun. ACM 55*, 11 (Nov. 2012).
21. Warnakulasooriya, R., Palazzo, D.J., and Pritchard, D.E.Time to completion of Web-based physics problems with tutoring.*Journal of the Experimental Analysis of Behavior 88*, 1 (2007).

丹尼尔 T·西顿（Daniel T. Seaton ）（dseaton@mit.edu ）是麻省理工学院数字化学习办公室的博士后研究员。

约阿夫·伯格纳（Yoav Bergner ）(ybergner@ets.org) 是新泽西州普林斯顿大学教育测试中心（ETS ）高级心理测验学中心的研究科学家。

艾萨克·庄（Isaac Chuang ）(ichuang@mit.edu) 同时担任麻省理工学院物理系及电子工程和计算机科学系的教授，也是麻省理工学院电子学研究实验室的成员。

彼得·密特罗斯（Piotr Mitros ）(piotr@mitros.org) 是麻省理工学院人工智能和学习中心下属 edX 的首席科学家。

戴维 E·普里查德（David E. Pritchard ）(dpritch@mit.edu) 是麻省理工学院物理学的 Cecil and Ida Green 教授，以及超冷原子中心和电子学研究实验室的成员。

译文责任编辑：谢涛

DOI:10.1145/2580723.2580730

AR 系统面临在它们普及之前就该解决的潜在安全问题。

作者 FRANZISKA ROESNER、TADAYOSHI KOHNO 和 DAVID MOLNAR

增强现实系统的安全和隐私

增强现实（AR）技术有望增强我们对现实世界的感知和交互。与用模拟世界取代现实世界的虚拟现实系统不同，AR 系统自动感知物理世界的属性，并实时将计算机生成的视觉、听觉和触觉信号覆盖在现实世界反馈上。本文中，我们考虑与 AR 系统本身有关的安全和隐私问题，及其支持技术产生的同类问题。

自上世纪 60 年代起，当 Ivan Sutherland 描绘显示三维信息的透明头戴式显示器时，研究人员就已经在探索 AR 的想法。[33] 从上世纪 90 年代起，AR 作为研究领域已将重点集中在克服显示技术、跟踪和配准方面的难题，以便正确对齐虚实对象、用户界面和人为因素、辅助传感设备，以及新型 AR 应用程序的设计。[1,2,6,22,36,41]

但是，直到最近，早期出现的 AR 技术才开始投入商用。例如，Google 最新推出了支持 AR 应用程序的有限数量的 Google Glass“平视”眼镜。很多其他早期 AR 应用程序也因智能手机和其他移动设备的普及而得到运用。这方面的例子包括 Word Lens iPhone 应用程序以及 Layar。前者将翻译后的文字叠加到相机的外语文字上；后者则是基于地理位置的 AR 平台，供开发者创建 AR 层，用于游戏等各种领域；请参阅图 1。最近在手机中出现的 1GHz 处理器、位置传感器以及高分辨率自动对焦相机已使这些应用成为可能。

本文中，我们将从宽泛的角度探讨 AR 领域，既考虑 AR 的直接应用，又考虑支持这些应用所必需的技术。除了手机之外，增强感觉、显示和数据共享的设备也已开始出现，它们将使更加复杂的 AR 系统成为现实。例如，Looxcie（一种耳挂式、始终开启的视频摄像头）具有使佩戴者能与世界上任何人共享实时视频源的功能。Microsoft 的 SDK for Kinect[20] 将 RGB 相机、深度相机和多阵列麦克风结合在一

» 重要见解

- **增强现实技术发展迅速，并且已投入商用，它们将给安全和隐私带来新的挑战和机遇。这些挑战的特征可沿两根轴来描述：系统范围和功能。**
- **AR 技术的安全和隐私挑战包括：共享输入和输出设备的应用程序之间的冲突，以及更为复杂的传感器数据访问控制。虽然某些问题可通过借鉴智能手机的现有解决方案来解决，但是其他问题还需要运用新的方法。**
- **AR 技术为以全新方式应对现有安全和隐私挑战提供了机遇。**

照片插图：BARRY DOWNARD


Keychain
Password:
ke1F367c22

图 1. 基于手机的增强现实。左侧是 Word Lens 的图片，这是一种提供无缝“图内”翻译的 iPhone 应用程序（来源：http://www.flickr.com/photos/neven/5269418871/）。在这里，应用程序将单词“craft”从英语翻成西班牙语，然后再译回英语。右侧是 Layar 的图片，这是 Android 手机搭载的“增强现实浏览器”（来源：http://site.layar.com/company/blog/make-your-ownlayar-screen-shot-with-the-dreamcatcher/）。

图 2. 可穿戴输入和输出。左侧是肯尼亚护林员佩戴的 Looxcie 可穿戴式摄像头（来源：http://looxcie.com/index.php/image-gallery）。右侧是 2012 年 6 月的 Google Glass 原型（来源：http://www.flickr.com/photos/azugaldia/7457645618）。

起来提供精准运动感知，实现了多种原型 AR 应用程序。除了 Google Glass 之外，Vuzix、Lumus 和 Meta SpaceGlasses 等多家公司出品的透明可穿戴显示屏现已可供研究之用。图 2 显示了此类输入和输出设备的示例。（表 1 提供了 AR 支持技术的总结；其中的很多技术已经投入应用，而其他技术仍处试验阶段。）

这些技术将使商用 AR 应用程序成为可能，它们正处于重大创新的风口浪尖，将令很多用户受益匪浅。但是，这些技术也带来了未曾意料的计算机安全和隐私风险。AR 领域前人的研究很少考虑这些问题。我们主张，与其坐等这些技术完全成熟后再回过头来开发安全和隐私保护措施，不如现在趁这些技术仍然年轻、有可塑性的时候，就考虑安全和隐私问题。为了引导此过程，我们提出以下问题：随着 AR 系统及其支持技术的兴起，安全和隐私研究面临哪些新的挑战？AR 技术为改进安全和隐私创造了哪些崭新的机遇？

我们发现 AR 技术为计算机安全和隐私研究及相关产业搭建了一个重要而机会良多的新平台。当然，这些技术也应充分利用标准的最佳安全方案，如设备加密和网络加密。尽管如此，我们仍发现了一些既考验智慧，但仍可克服的独特障碍，包括如何处理共享一个 AR 系统输出的多个应用程序之间的冲突。其他挑战（如数据访问控制）在其他领域已经众所周知，但对于输入始终开启、始终感测的 AR 技术来说，这些挑战有着更重要的意义。鉴于 AR 技术在未来的重要性，在其他领域已经克服这些问题的研究者会发现，将注意力重新集中到 AR 应用具有重要价值。

除了提出新的挑战外，AR 系统还为可改进安全和隐私的新应用带来了机遇。例如，这些技术可在个人显示屏上提供个人的数字化内容视图。想象一下，一款密码管理器可在用户注视键盘时，将视觉的指示器叠放在某个复杂密码的正确键上，或者一款应用程序可在某人说谎时提醒用户。

本文中，我们将探讨 AR 技术带来的安全和隐私挑战、防御方向，以及 AR 系统对现有安全和隐私问题的新应用。

挑战

除了传统上定义的实时配准虚实对象外，我们考虑的 AR 应用程序和技术可能具有以下任一或全部特征：

- 始终开启的输入设备和传感器的复杂集合（例如相机、GPS、麦克风）。
- 多个输出设备（例如，显示屏、耳机）。
- 可同时运行多种应用程序的平台。
- 通过无线方式与其他 AR 系统进行通信的能力。

这里，我们提出一系列由这些新技术及其应用带来的安全和隐私挑战（已在表 2 中总结）。我们将这些挑战沿两根轴来组织：*系统适用范围*和*功能*。在一根轴上，我们考虑适用范围逐渐扩大的 AR 系统：从单一应用程序到单个 AR 平台中的多个应用程序，再到多个相互通信的 AR 系统。各类别的挑战在系统复杂度达到这一程度时首次出现。对于每个范围，我们进一步将挑战划分为与输入、输出或数据访问相关等不同类别。我们建议以后

的 AR 技术设计者沿这两根轴考虑安全和隐私挑战。

熟悉智能手机安全性的读者可能会发现，手机安全挑战与我们这里所提出挑战之间有某种重叠。我们注意到，某些智能手机安全技术可能适用于 AR 技术；其他技术需要在此新背景下重新考虑。

单一应用程序情况下的挑战。我们首先仅考虑单一 AR 应用程序的威胁和挑战。

*输出。*用户必须十分信任那些将虚拟反馈叠加于真实视觉、听觉或触觉感知之上的 AR 应用程序。提供沉浸式反馈的设备可能被恶意应用程序利用来欺骗用户，使其错误认识真实世界。例如，未来的恶意应用程序可能在真实限速标志上叠放不正确的限速标志（或者在没有标志的地方放置虚假标志），或者故意提供真实外语文本的错误译文。更笼统地说，这样的应用程序可能欺骗用户，使其错误地认为某些物体在现实世界中存在或不存在。

恶意应用程序可使用类似的技术造成用户感官超载。应用程序在屏幕上闪烁亮光、播放巨响的声音，或者产生剧烈的触觉反馈，这些都可能对用户造成人身伤害。这样的攻击并非没有先例：攻击者曾针对癫痫论坛发布动态 gif 闪光图，结果引起浏览者头疼或癫痫发作。[24] 新兴的 AR 平台必须考虑并防止这些类型的攻击。

这些输出攻击一旦出现在沉浸式 AR 应用程序中，带来的严重后果将远胜在目前的桌面或手持式计算场景中带来的后果，这既因为用户更难区分虚实反馈，也因为用户可能更难移除或关闭系统。作为对付输出攻击的最后手段，用户必须能够轻松可靠地返回到现实世界，也就是说，可以确认所有输出设备已被关闭。

近期来看，移除系统是实现这种现实回归的简单方法。但是，未来的可穿戴系统可能很难甚至不可能让用户移除（例如，隐形眼镜 [23] 或植入式设备），今天的不可穿戴系统可能已经让用户难以摆脱。例如，多家汽车制造商生产了能在用户道路视野上显示增强内容的挡风玻璃。[5] 在这些案例中，系统应该有可让用户返回现实的可信途径，类似于 Windows 计算机上的 Ctrl-Alt-Del。要确定此类最佳回归顺序，或者说正确的输入模式（例如，手势或语音），需要对每种 AR 系统进行研究。另一种方法可能是在显示屏上保留一块总是显示现实世界的可信区域。

*输入。*AR 应用程序无疑将面临与传统应用程序相似的输入验证和去害难题。例如，解析现实中文字的翻译应用程序可能被某指示牌上恶意制作的文字不当利用。传统输入验证技术可能仍适用，但 AR 系统的设计者应意识到它们在这种新情况下的必要性。

*数据访问。*为了提供预期的功能，AR 应用程序可能需要访问各种传感器数据，包括视频和音频、GPS 数据、温度、加速度计读数等。与桌面和智能手机操作系统一样，AR 系统的重要挑战将是，在实现功能所需要的访问与应用程序窃取数据或滥用这种访问的风险之间做出权衡。例如，恶意应用程序可能向其后台服务器泄露用户位置或视频。现有的概念验证性 PlaceRaider 攻击 [34] 显示，智能手机传感器可用来收集足够的信息以创建室内环境的三维模型。

与当前大多数桌面和智能手机应用程序不同的是，复杂的 AR 应用程序将需要丰富、始终开启的传感器。例如，自动检测并扫描 QR 码的应用程序需要持续访问视频流数据，而自动检测用户何时在另一台设备上输入密码并提供密码帮助的应用程序也同样如此（我们接下来将对此进行讨论）。因此，这些隐私风险比传统系统大得多。

表 1. 商用和新兴 AR 技术总结。

	目前已投入商用	仅为试验性
传感器（输入）	可穿戴式 RGB 相机 GPS（误差 5 米或更多）精确动作感应（例如，Kinect）	触觉传感器 [29]
反馈（输出）	不透明近目显示屏手机显示屏 / 扬声器不可见蓝牙耳机	透明近目显示屏嵌入式显示器（隐形眼镜 [23]）触觉反馈 [17]
服务	简单云服务（照片库）基于标记的跟踪 [39] 良好面部检测（非识别）[37] 昂贵或廉价但不精确的转录	复杂云服务（对象识别）无需标记物的跟踪良好面部识别廉价但精确的转录
共享	选择性共享（照片、视频、位置）	自动共享

表 2.AR 技术的安全和隐私挑战。我们按两根轴给这些挑战分类：与输出、输入和数据访问相关的挑战，以及出现在单一应用程序、多应用程序系统和多个相互交互的系统中的挑战。

	单一应用程序	多应用程序	多系统
输出	欺骗攻击 过载攻击 回归现实的可信途径	处理冲突 点击劫持	相互冲突的视图
输入	输入验证	解析焦点	聚合输入
数据访问	传感器数据的访问控制 旁观者隐私	跨应用程序共享	跨系统共享

AR 系统应采取可控制这些风险的方法。例如，个别应用程序可能不需要访问所有传感器数据。当用户处于某个位置时，也许应用程序只需要访问部分屏幕，或者只需要知道系统识别的某些物体（例如，通过 Kinect 的骨架识别器），而无需访问全部原始相机输入。AR 系统设计者必须考虑这些权限的适当粒度，而且易用的权限管理界面的设计将非常重要。智能手机中使用的基于清单或提示的现有解决方案不太可能以有效的方式扩展；由于 AR 应用程序需要长期（而非一次性）访问数据，这使得上下文内访问控制解决方案（如用户主导的访问控制）的应用 [28] 不那么简单。

始终开启的相机和其他传感器还会给旁观者带来隐私风险，Krevelen 和 Poelman 认为这会影响 AR 得到社会广泛接受。[36] 旁观者应该能在他人录像时选择避开或隐匿身份（例如，模糊影像）；以前的研究曾考察过此类问题。[9,31]AR 用户可能需要某些方式来向心存疑虑的旁观者证明这样的防护措施已然就位。立法或市场力量可能催生出对来自其他设备或环境的请求作出响应的相机；新闻报道披露，Apple 已经考虑将这样的功能添加到 iPhone 来防止偷拍诸如音乐会之类的现场活动。[4] 相机还可在录制时提醒旁观者，例如通过闪光 [36] 或提供访问更复杂的政策信息的权限。[19]

CVDazzle 项目 [10] 另辟蹊径，使用化妆来迷惑面部检测算法 — 这种方法无需隐私保护相机即可提供私密性。CVDazzle 的关键局限性是需要费力地为一种特定的面部检测算法进行手工调优。这里的一个研究课题是，如何找到一种通用算法来合成可迷惑面部检测的化妆。

多应用程序挑战。虽然 AR 应用程序通常是独立构想并开发原型的，但是我们可以预期，未来的 AR 平台（如基于 Google Glass 或 Microsoft Kinect 构建的平台）将支持同时运行的多个应用程序，这些应用程序共享输入和输出设备，并相互公开数据和 API（请见图 3）。研究人员必须预见这些发展趋势，并确保在设计“支持增强现实的操作系统”时，适当考虑安全和隐私。

*输出。*在多应用程序 AR 系统中，各个应用程序将共享输出设备，包括显示屏、音频输出和触觉反馈。试图使用这些输出设备的多个应用程序之间若发生冲突，可能导致安全问题。例如，恶意应用程序可能试图遮蔽另一个应用程序呈现的内容（例如，在视觉或听觉上用不正确的翻译掩盖正确的翻译）。

图 3. 多应用程序 AR。新兴及未来的 AR 平台将支持同时运行的多个应用程序，这些应用程序共享输入和输出设备，并相互公开数据和 API。在多应用程序 AR 系统中，各个应用程序（正如此模型中所描绘的应用程序）将共享输出设备，包括显示屏、音频输出和触觉反馈。这些应用程序之间的冲突可能导致安全问题。

尽管如此，为了在 AR 系统中提供所需的功能，输出共享仍是必需的。例如，用户可能希望同时查看覆盖在实景视图上的来自多个应用程序的内容，例如，地图应用程序提供的方向、汇总邻近好友活动的社交信息、音乐应用程序当前播放的曲目等等。因此，一次只有一个应用程序控制显示屏的原生解决方案并不够用。

因此，未来的 AR 系统必须处理多个尝试生成输出的应用程序之间发生的冲突。例如，五个应用程序可能全都要标注同一个对象（例如，使用翻译字幕），系统需要为它们排列优先级。此外，用户应能知道哪些内容是由哪个应用程序生成的，这一点可能也很重要 — 例如，标注的产品推荐是来自好友还是广告商。AR 系统设计者所创建的界面必须可让用户清楚知道或轻松发现所示内容的来源。

基于输出篡改的传统攻击在 AR 环境下可能需要新的方法或新的规划。例如，在目前的系统中，应用程序可能发动点击劫持攻击，这种攻击诱骗用户点击另一个应用程序中的敏感用户界面元素（例如，在用户社交媒体档案中发布某些内容）。发动这些攻击的方式通常是，篡改敏感元素的显示（使其透明或以巧妙的方式将其部分遮盖），或者就在用户点击可预测的位置之前突然显示敏感元素。未来 AR 系统的应用可能开发出诱骗用户与元素进行交互的新技术，而系统设计者必须预见到这些攻击。例如，AR 应用程序可能试图诱骗用户与现实世界（而非虚拟世界）中的对象交互。

*输入。*用户可能不会使用传统输入方法（如点击鼠标或使用触摸屏）来与 AR 系统交互。相反，用户可能越来越多地使用触觉传感器（例如，嵌在手套中）的精细输入、使用语音，或在视线跟踪技术的帮

助下，来与系统交互。有了这些输入技术和多个运行的应用程序，让系统解析哪个应用程序处于焦点位置、并因此而应接收输入就显得非常重要。

例如，目前的语音交互发生在用户以显式地动作指明了目标应用程序之后（例如，单击 iPhone 上的“Siri”按钮），或者发生在只有一个应用程序可以接收语音输入的系统上（例如，在 Xbox 上）。当多个应用程序处于活动状态，并可能在任何给定时间接收语音或其他输入时，必须有一种有效的方法能让用户使应用程序获得焦点，或者在焦点不明的情况下，让系统决定输入命令发给哪个正确目标。我们强调：未来的 AR 系统很可能同时运行多个应用程序，其中的很多应用程序将一直运行并侦听输入，而没有任何可见的输出。设计不当的焦点程序解析可能让恶意应用程序轻松窃取本该发给另一个应用程序的用户输入（例如，窃取本该发给另一应用程序的登录框的密码）。例如，恶意应用程序可能尝试注册与另一个敏感的应用程序所使用的发音相似的口头关键字，从而有意增加输入歧义性。

*数据访问。*跟传统操作系统一样，AR 应用程序可能希望相互公开 API，而用户可能希望在应用程序之间共享虚拟对象。研究人员必须针对跨应用程序共享摸索出恰当的访问控制模型。从传统访问控制设计得到的某些经验教训可能在此领域中适用，但是新的技术和环境可能需要新的方法。例如，复制 - 黏贴和拖放是长期固定下来的在传统应用程序之间共享数据的用户手势，因此在访问控制方面仍有意义。桌面和智能手机系统领域的大量研究已尝试将用户操作与应用程序特权对应起来（例如 Miller[21] 和 Roesner 等人 [28] 的研究）；AR 系统需要发展出新的用户手势来表明共享意图。此外，AR 系统不太可能像传统桌面操作系统那样，在带标签的窗口中显示应用程序，因此我们需要新的交互范式来使用户能识别应用程序，并指出哪个应用程序应该接收共享数据。

我们主张趁现在这些技术仍然年轻、有可塑性的时候，就考虑 AR 安全和隐私问题。

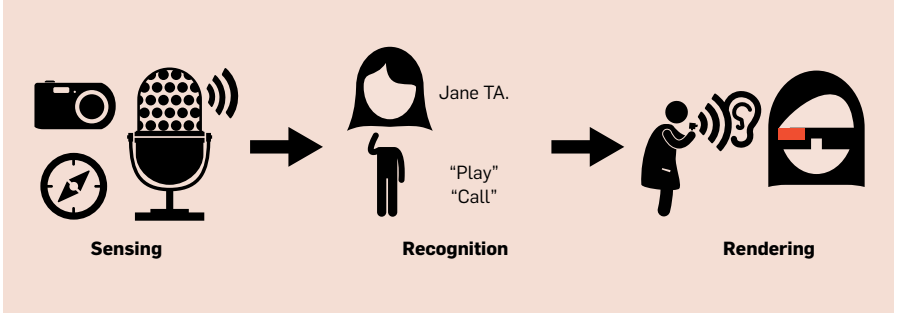
多系统挑战。跳出运行多个应用程序的单个 AR 系统，我们将考虑属于不同用户的多个 AR 系统间的交互。以前的 AR 研究提出了在一个 AR 系统的多用户之间协作的应用程序。这些应用程序包括多人游戏、[11,32,40] 远程会议现场遥现 [16] 以及面对面协作。[26] 这类应用程序带来了更多安全和隐私挑战。

*输出。*不同的用户可能看到由其各自 AR 系统所呈现的不同图景。例如，不同的用户可能看到叠加在现实公告牌上的不同的虚拟广告，或者根据用户的访问权限级别，可能向观看演示的不同用户显示不同的内容（例如，一个用户可能看到绝密脚注，而且其他用户则不然）。如此相互冲突的视图要求用户妥善管理关于“谁可以感知哪些信息”的思维模式，以免他们意外透露只有自己才可使用的私密信息。要解决此问题，需要在界面设计方面开展创新，以便协助用户完成此类任务。

*输入。*当支持技术提供的传感器输入在数量和复杂性方面呈上升态势时，AR 系统和应用程序的复杂性也会随之上升，两者是密切相关的。如此大量的来自众多用户的传感器输入反过来会催生新的协作式传感应用程序，这些应用程序本身可向 AR 应用程序反馈数据。例如，Google 已经在使用由用户智能手机收集的数据来估计交通路况，然后报告给用户的手机。[8] 要实现未来可显示在汽车挡风玻璃上的 AR 应用程序，这类数据是必需的。

但是，这类聚合输入可被恶意用户用来愚弄数据收集系统。例如，评论网站可能利用位置跟踪，通过

图 4.AR 流程。AR 应用程序 (1) 收集传感数据，它们 (2) 使用高级语义从这些数据中提取对象。最后，它们 (3) 在用户感觉的基础上呈现内容。



标注当天到场的平均人数来评测餐馆的热门程度。精明的餐馆老板到时可能出钱请人到餐馆站场，而这些人不买任何东西。餐馆的评测热门度会上升，但跟其服务质量毫无关系。

不断收集数据的 AR 技术将推动此类协作感测应用程序的普及；因此，这些安全问题的重要性也会上升。另举一个例子，社区地震网络聚合了很多个人的测震仪传感器数据来检测和预测地震；攻击者可能操纵这些传感器来“伪造”异常地震活动，例如，鼓动受此项目监测的很多人在另外一个毫不相干的游戏中同时跳起。（例如，Improv Everything[13] 要求用户在指定的时间播放提供的音频文件，并按照音频指示行动。）可信任的传感器 [30] 虽然对防止其他攻击非常重要，但在这种情况下毫无作用，因为现实情况受人操控。

*数据访问。*除了向不同的用户显示不同的内容外，相互通信的 AR 系统将允许用户相互共享虚拟内容。例如，一个用户可能在其私有 AR 系统中创建虚拟文档，然后选择与其他用户的系统共享其显示内容。某些共享甚至可能很隐秘；想象一下，某个 AR 系统自动使用邻近用户的相机源来向某个给定用户提供他或她的实时 3D 模型。

跨不同 AR 系统的隐式或显式共享数据可使很多有价值的应用成为现实。但是，这需要相应的访问控制模型和界面来允许用户管理这种共享。现在，由于人与数据项之间的复杂关系，用户已经很难针对 Facebook 等服务上的隐私设置形成自己的思维模式。[18]AR 系统收集的大量数据以及虚拟对象与现实世界的集成只会使这个问题更加严峻。

防御方向

这里，我们将概述 AR 技术的多个防御方向。首先，与 AR 技术相关的某些安全和隐私挑战与目前智能手机所面临的那些挑战（例如，传感器数据的隐私和跨应用程序共享）十分相似。某些情况下，恰当的 AR 的防御方向是借鉴并相应调整智能手机解决方案。例如，可以在短期内采用许可清单和应用商店审查流程。

但长期来看，有多种原因造成 AR 环境下的方法必须有别于智能手机解决方案。首先，对智能手机应用程序的资源需求分析 [28] 显示，大多数人只需要一次性或在短期内访问大多数资源，这使得需要上下文内用户交互的解决方案（如用户主导的访问控制 [28]）具有可行性。相比之下，AR 应用程序需要长期或永久访问传感器数据，而且在规模上超越了智能手机应用程序。此外，AR 资源访问对用户和旁观者来说，不如在智能手机环境中那样明确 — 例如，AR 系统的相机将始终开启，而智能手机的相机即使被恶意软件开启，当它放在用户口袋里的时候，也不会提供很多数据。因此，我们认为在这一领域设计解决方案时，应全面考虑未来的 AR 环境。

除此之外，在 AR 专用解决方案方面还需要开展新的研究。例如，研究人员已经开始考虑 AR 特有的操作系统支持。[7]AR 应用程序以及底层操作系统自然地遵循图 4 所示的流程，因此我们可以相应地确定研究方向，并且不同的研究模型可以在应用程序和操作系统之间假设不同的边界。在第一阶段“传感”中，应用程序（或操作系统）收集原始传感数据，如音频、视频或无线电波；这里的研究包括限制收集哪些传感信息（例如，“礼貌”相机 [9,31]）或限制这些信息的使用（例如，保留策略）。其次，在识别阶段，机器学习算法通过高级语义提取对象：例如，图中显示的 Kinect 骨架、面部、关联的姓名和语音命令触发器。相关研究包括更改对象以造成漏报（例如，CVDazzle[10]），以及支配应用程序访问对象的策略。[15] 最后，应用程序（或操作系统）在用户感觉的基础上呈现视觉和听觉等内容。这里的研究包括：发现那些为避免伤害用户而必须遵守的约束条件，以及构建遵守这些约束条件的高性能“可信任呈现器”。

并非所有 AR 防御方向都会由技术解决方案组成。某些挑战可能需要社交、策略或法律方法；例如，前面讨论过的旁观者屏蔽和隐私保护相机的潜在策略。同样，其他问题也将从非技术方法中获益。

最后，我们呼吁为在此领域工作的研究者提供 AR 测试台。今天的大多数试验性 AR 应用程序依赖于 Microsoft Kinect 或 Layar 之类的智能手机平台；两者只涉及一次性运行的单一应用程序，因而隐藏了随着 AR 系统的复杂性增加而出现的挑战。

新应用程序

虽然 AR 技术产生了重要的安全和隐私问题，但通过将这些技术运用到现有问题上，它们有机会增强安全和隐私，只是这样的机会目前尚未充分利用。这里，我们考虑了由 AR 技术和系统实现的新型安全和隐私增强应用程序所带来的机遇。我们的列表无疑还不完整；我们希望未来在此领域看到丰硕的成果。

利用个人视野。集成平视式或其他个人显示屏的 AR 系统可利用这种个人视野来解决现有的安全和隐私问题 — 尤其是保护私有数据和改进密码管理。

个人显示屏可以很好地防止肩窥，因为用户可与只在自己的视野中可见的应用程序交互。例如，在现在的飞机上使用笔记本电脑的人会把他看到和输入的所有内容暴露给邻座的人，研究人员已经证明，低成本相机拍摄的视频即可重构用户在虚拟移动键盘上的键入。[25] 个人平视式显示屏若与用于隐蔽输入的触觉传感器相结合，将大大提高隐私性。[a]

个人显示屏可在现实世界中进一步增强内容加密，使得只有预期接收人的 AR 系统可以解密。例如，公司可以在公告牌上发布加密通知，员工可通过公司配发的 AR 系统阅读这些通知，但是公司大楼的访客无法阅读。（在 AR 系统可访问的服务器上只存储密钥，而不是加密内容，这样竞争对手就只能去寻找实体通知，而不是攻破公司服务器即可了事。）此类系统的前身如今已借助智能手机和 2D 条形码（将 URL 编码为具有相应访问控制的数据）实现；增强平视式显示屏将省去手动扫描的需要。

AR 系统还可充当用户的增强密码管理器，通过个人显示屏显示密码或密码提示。例如，显示屏可以概略显示用户在老式设备（例如 ATM PIN 键盘）上必须输入的相应字符。届时，就可以为用户分配强密码，因为他们无需实际记住密码。这种应用离不开无痕跟踪以及可妥善保护所存储密码的系统设计。

举一个具体示例，我们实现了由 Google Glass 应用程序和浏览器 (Chrome) 扩展组成的原型密码管理器应用程序（请参阅图 5）。Chrome 扩展修改了浏览器的用户界面，以显示代表当前所示网站的 QR 代码（浏览器地址栏中的网站）。用户可要求 Google Glass 应用程序扫描这些 QR 代码，并使用语音命令“OK Glass, find password”来查询密码数据库。如果用户以前存储过该网站的密码，则应用程序将显示密码；否则，用户可登记新密码，他们可以要求 Chrome 扩展生成登记 QR 码，并使用“enroll password”语音命令存储新密码。我们已经在 https://github.com/froeschele/GlassPass 上公开了我们原型的代码。

在设计由浏览器扩展显示的 QR 码时，我们纳入了浏览器和手机共享的机密，这样一来，此应用程序还可提供钓鱼保护，因为网站无法创建并显示映射到密码管理器中的合法密码的伪造 QR 码。

利用复杂传感器系统。AR 系统得益于多种输入和传感设备的组合，这些设备结合在一起，可增强数字及物理安全性和隐私性。

未来的系统可利用 AR 技术来*检测应提醒用户注意的隐私或安全条件*。例如，系统可以在检测到相机镜头对准用户时提醒用户（例如，使用计算机视觉来检测镜头发射出的闪光），而不是依赖隐私保护相机来将用户从不需要的录像中屏蔽。[35] 还可以检测某些形式的窃听，例如，指向窗口的激光麦克风。

这样的系统还可*检测物理欺骗企图*。例如，AR 系统可以估计 ATM 卡槽的大小和形状，如果发现

图 5. 原型 AR 密码管理器。我们的 Chrome 扩展（背景）显示了表示当前网站的 QR 码。响应语音命令“find password”时，我们的 Google Glass 应用程序（前景）扫描此 QR 码，并在平视式显示屏上私密地显示为该网站存储的密码。

a 我们注意到，从外部观察者的角度，透视显示屏（如 Google Glass 所用的显示屏）并非完全私密。例如，类似于根据屏幕反射重构内容，使用远距镜头拍摄的显示屏图像可用来重构屏幕内容。[3] 未来的研究应充分揭示此类威胁的特征，并设计相应的防御措施。

似乎加装了盗刷装置，系统就会发出警告。同样，面部表情自动识别方面的现有研究[12]也可运用于基于行为的谎言检测。[38]我们的某位同事将此应用称为“蜘蛛侠感觉”。

除了存储密码外，AR 系统还可用于隐式鉴别用户身份。使用这些技术以及附着在人身上的大量传感器都可用来根据生物特征和行为特征鉴定用户身份。以前的研究考察了在手机上实现此类机制的可能性。[14,27]AR 系统将提供更为强大的身份验证功能。传感器数据同样有助于做出授权和访问控制决策。

除了附加在个人（比方说 Alice）身上的传感器外，旁观者的传感器也可用来验证她的身份（将 Alice 的第三方视觉、听觉和其他感官视图提供给身份验证系统）。此第三方身份验证系统将信任那些没有动机以虚假方式验证 Alice 的系统和人员。

总结

AR 系统具有精密而普适的输入、输出和处理能力，它们具备让很多用户明显受益的潜力。为了促进 AR 技术的持续创新，我们认为，在 AR 系统得到广泛部署，并且其体系结构得以确定之前，现在还应制定路线图来保护 AR 系统的计算机安全和个人隐私。为了催生此路线图，我们考虑了这些系统所面临的新的安全和隐私挑战，并且探索了由这些技术带来的机遇，希望籍此创建新的隐私和安全得到增强的应用程序。

鸣谢

本文部分得到美国国家科学基金会（资助项目 CNS-0846065、CNS-0905384，及资助项目 DGE-0718124 名下的研究生研究奖学金计划）以及微软学者奖学金项目的支持。我们向 Luis Ceze、Lydia Chilton、Alexei Czeskis、Nicki Dell、Tamara Denning、Karl Koscher、Brandon Lucia、Alex Moshchuk、Bryan Parno、Karin Strauss Helen Wang 以及匿名审阅者表示衷心感谢。

参考资料

1. Azuma, R.T.A survey of augmented reality. *Presence:Teleoperators and Virtual Environments 6* (1997), 355–385.
2. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and Macintyre, B. Recent advances in augmented reality.*IEEE Computer Graphics and Applications 21*, 6 (2001), 34–47.
3. Backes, M., Chen, T., Duermuth, M., Lensch, H. and Welk, M. Tempest in a teapot:Compromising reflections revisited.*IEEE Symposium on Security and Privacy* (2009).
4. Business Insider.This apple patent will shut down your camera at live concerts; http://www.businessinsider.com/iphone-concert-patent-2011-6.
5. CNN.Augmented-reality windshields and the future of driving, 2012; http://virtual.vtt.fi/virtual/proj2/multimedia/alvar.html.
6. Costanza, E., Kunz, A. and Fjeld, M. Mixed reality:A survey.*Human Machine Interaction.* Springer-Verlag, 2009, 47–68.
7. D’Antoni, L., Dunn, A., Jana, S., et al.Operating system support for augmented reality applications. In *Proceedings of USENIX Workshop on Hot Topics in Operating Systems* (2013).
8. Google.Crowdsourcing road congestion data; http://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html.
9. Halderman, J.A., Waters, B. and Felten, E.W.Privacy management for portable recording devices.In *Proceedings of the 3rd ACM Workshop on Privacy in Electronic Society* (2004).
10. Harvey, A. CVDazzle:Camouflage from Computer Vision; http://cvdazzle.com/.
11. Henrysson, A., Billinghurst, M., And Ollila, M. Face to face collaborative AR on mobile phones.In *Proceeding of the 4th IEEE/ACM International Symposium on Mixed & Augmented Reality* (2005).
12. Hoque, M.E., McDuff, D. and Picard, R.W.Exploring temporal patterns in classifying frustrated and delighted smiles.*IEEE Transactions on Affective Computing 3* (2012), 323–334.
13. Improv Everywhere.The Mp3 Experiments, 2012; http://improveverywhere.com/missions/the-mp3-experiments/.
14. Jakobsson, M., Shi, E., Golle, P., and Chow, R. Implicit authentication for mobile devices.In *Proceedings of the 4th USENIX Workshop on Hot Topics in Security* (2009), USENIX.
15. Jana, S., Molnar, D., Moshchuk, A. et al.Enabling fine-grained permissions for augmented reality applications with recognizers.Tech.Rep. MSR-TR-2013-11, Microsoft Research, Feb. 2013.
16. Kato, H. and Billinghurst, M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system.In *IEEE/ACM Workshop on Augmented Reality* (1999).
17. Laycock, S. and Day, A. A survey of haptic rendering techniques.*Comp.Graphics Forum.26*, 1 (2007), 50–65.
18. Madejski, M., Johnson, M. and Bellovin, S.M.The Failure of Online Social Network Privacy Settings.Tech.Rep. CUCS-010-11, Dept. of Comp.Science, Columbia University, 2011.
19. Maganis, G., Jung, J., Kohno, T. et al.Sensor Tricorder:What does that sensor know about me?In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications* (2011), ACM.
20. Microsoft.Kinect for Windows, 2012; http://www.microsoft.com/en-us/kinectforwindows/.
21. Miller, M.S.Robust Composition:Towards a Unified Approach to Access Control and Concurrency Control. Ph.D. thesis.Johns Hopkins University, Baltimore, MD, 2006.
22. Papagiannakis, G., Singh, G. and Magnenatthalmann, N. A survey of mobile and wireless technologies for augmented reality systems.*Computer Animation and Virtual Worlds 19* (2008), 3–22.
23. Parviz, B. For your eye only.*IEEE Spectrum 46* (2009), 36–41.
24. Poulsen, K. Hackers assault epilepsy patients via computer.*Wired* (2008); http://www.wired.com/politics/security/news/2008/03/epilepsy.
25. Raguram, R., White, A.M., Goswami, D. et al. iSpy: automatic reconstruction of typed input from compromising reflections.In *Proceedings of the 18th ACM Conf.Computer and Communications Security.*
26. Reitmayr, G. and Schmalstieg, D. Mobile collaborative augmented reality.In *Proceedings of the 4th International Symp. on Augmented Reality* (2001).
27. Riva, O., Qin, C., Strauss, K., And Lymberopoulos, D. Progressive authentication:Deciding when to authenticate on mobile phones.In *Proceedings of the 21st USENIX Security Symposium* (2012).
28. Roesner, F., Kohno, T., Moshchuk, A. et al.User-driven access control:Rethinking permission granting in modern operating systems.*IEEE Symposium on Security and Privacy* (2012).
29. Saponas, T.S., Tan, D.S., Morris, D. et al.Enabling always-available input with muscle-computer interfaces.In *Proceedings of the 22nd ACM Symposium on User Interface Software and Technology* (2009).
30. Saroiu, S. and Wolman, A. I am a sensor, and I approve this message.In *Proceedings of the 11th Workshop on Mobile Computing Systems and Applications* (2010), ACM.
31. Schiff, J., Meingast, M., Mulligan, D.K., Sastry, S. and Goldberg, K.Y.Respectful cameras:Detecting visual markers in real-time to address privacy concerns.In *Proceeding of the 2007 Int'l Conference on Intelligent Robots and Systems.*
32. Starner, T., Leibe, B., Singletary, B. and Pair, J. Mindwarping:Towards creating a compelling collaborative augmented reality game.In *ACM Intelligent User Interfaces* (2000).
33. Sutherland, I.E.A head-mounted three-dimensional display.In *Proceedings of the Fall Joint Computer Conference, American Federation of Information Processing Societies* (1968).
34. Templeman, R., Rahman, Z., Crandall, D.J., And Kapadia, A. Placeraider:Virtual theft in physical spaces with smartphones.CoRR abs/1209.5982 (2012).
35. Truong, K., Patel, S., Summet, J. and Abowd, G. Preventing camera recording by designing a capture-resistant environment.*Proceedings of Ubicomp* (2005).
36. Van Krevelen, D. and Poelman, R. A survey of augmented reality technologies, applications, and limitations.*The International Journal of Virtual Reality 9* (2010), 1–20.
37. Viola, P. and Jones, M. Robust real-time object detection.*International Journal of Computer Vision 57* (2004), 137–154, Hingham, MA.
38. Vrij, A., Edward, K., Roberts, K. and Bull, R. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior 24* (2000), 239–263.
39. VTT Technical Research Centre Of Finland.Alvar Software Library, 2009. http://cnn.com/2012/01/13/tech/innovation/ces-future-driving/.
40. Wagner, D., Pintaric, T., Ledermann, F. and Schmalstieg, D. Towards massively multi-user augmented reality on handheld devices.In *Proceedings of the 3rd International Conference on Pervasive Computing* (2005).
41. Zhou, F., Duh, H. B.-L. and Billinghurst, M. Trends in augmented reality tracking, interaction and display:A review of 10 years of ISMAR.In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (2008).

Franziska Roesner (franzi@cs.washington.edu) 是华盛顿大学西雅图分校的博士生。

Tadayoshi Kohno (yoshi@cs.washington.edu) 是华盛顿大学西雅图分校的副教授。

David Molnar (dmolnar@microsoft.com) 是华盛顿州雷蒙德 Microsoft Research 的研究人员。

译文责任编辑：山世光

研究亮点

技术视角
形变方法的“合理”解决方案

如需查看相应的论文，
请访问 /10.1145/2578851

作者：Joe Warren

几何在现代计算领域起着举足轻重的作用。在科学和工程学中，几何的数学模型对仿真和制造等应用至关重要。在艺术和娱乐领域，几何的数学模型在游戏和电影等应用中无处不在，甚至对图像编辑也非常有用。开发这些模型既直观又计算高效的新变体一直是计算机图形学领域的研究热点。

这个领域的一个经典问题是关于可变形建模；即，通过某种数学方法（理想情况下是交互式方法）将给定形状变形为目标形状。此问题的解决方案是大多数当前计算机动画系统的核心。这些方法的一种标准方式是，将形状视作由可变形的材料组成，并有一组嵌入的形状图柄。通常，这些形状图柄是点或者是连接这些点的分段线性形状，用户可交互式处理这些形状。在此框架下，形变方法根据某种简单的物理模型计算形状的“合理”形变。

这里的一个有意思的问题是，与“合理”形变相对应的是什么。从历史上看，最简单的形变数学模型一度是分段多项式。在一维的情况下，这些模型（如 B-spline）简单、直观、易于表现。对于二维形状，分段线性形变同样简单直观。但是，更高阶分段多项式模型缺乏很多应用所需要的平滑度和灵活性。由于一元多项式是简单微分方程的解，因此很多现代形变方案将平滑形变建模为偏微分方程 (PDE) 的解。基于 PDE 的大多数形变方法通常侧重于调和函数（拉普拉斯方程的解），或者最近的双调和函数（迭代拉普拉斯方程的解）。

在这方面，Jacobson 等人构造了一种形变方法，这种方法容许各种图柄类型（点、线段、开放和闭合多边形），并产生作为双调和函数的形变。在此框架下，图柄的位置被视作相关 PDE 的边界条件。要注意的一个重要事实是，取代调和函数而使用双调和函数是因为包含了孤立图柄。调和函数通常用来建模弹性薄膜的理想化形变，而双调和函数通常用来建模弹性薄板的形变。将孤立的图柄与薄膜（调和）内插将导致非平滑形变（通常自身折叠）。而将孤立的图柄与薄板（双调和）内插通常产生无折叠的平滑形变。

Jacobson 等人的主要技术创新是，为其模型添加了线性不等式约束，以确保得到的调和方程解是有界的。基于偏微分方程的大多数建模方法将自身限制在线性约束内，以确保求解过程是交互级别的。这些纯线性方法的缺点是，非负边界值问题的可能偶尔求解出到处不是严格非负的解。实际上，此问题使得将所需形变建模为通过一次干扰一个形状图柄而形成的形变组合变得更不直观。

此问题的数学解决方案是使用线性等式和不等式约束的组合来确保得到的 PDE 解非负。使用作者对这些约束的公式，产生的问题有界且凸起，因此可使用标准的稀疏二次规划求解器来有效求解。实际上，该方法离散化形状，并为每个单独的图柄预先计算独立的形变。这一主要求解过程作为预先计算完成，并且大多数案例只要几秒到十几秒。由于问题是线性的，因此采用每个独立图柄预先计算的形变的线性组合，可以交互式地计算出所需的对应于一组特定图柄位置的形变。局部性和凸性确保得到的形变直观地遵循边界条件。

这种方法是形变方法最高水平的极佳例证。该方法结合使用双调和方程来实现平滑的局部形变，同时使用线性不等式约束来将这些形变限制为非负且不存在局部极小值。由于形变只在形状上被计算为 PDE 的解，因此产生的形变只取决于形状中在距离上靠近的图柄位置。更广泛地说，该论文为使用高级数学方法来呈现复杂形变以及在实践中使用复杂数值求解器来计算这些形变的有趣的未来研究指明了方向。

Joe Warren (jwarren@rice.edu

译文责任编辑：周昆

DOI:10.1145/2578850

面向实时形变的有界双调和权重

作者：Alec Jacobson、Ilya Baran、Jovan Popovic 和 Olga Sorkine-Hornung

摘要

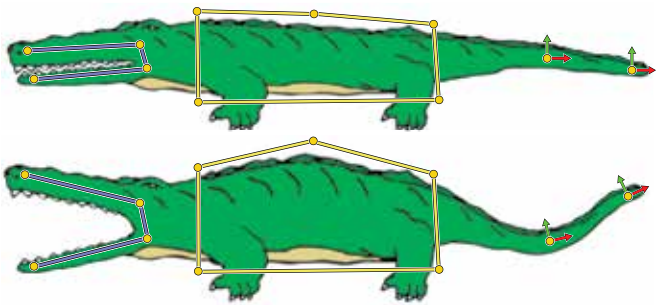
更改物体形状是计算机图形学中的一种基本操作，是变换光栅图像、矢量图形、几何模型和动画人物所必需的。这种物体形变中最快的方法之一是线性混合少量给定的仿射变换，通常每个变换分别与内部骨架的骨骼、封闭罩 (cage) 的顶点或一组零散的点图柄相关联。遗憾的是，线性混合方案往往不易于使用，因为它们可能需要手动绘制影响力权重或建立封闭多面体罩来包围输入物体。我们的目标是允许用户自由地对各类图柄进行最为方便的组合使用，从而让形变的设计和控制更加简单。我们开发了可对任意拓扑的点、骨骼和罩产生平滑直观形变的线性混合权重。我们的权重称为有界双调和权重，它在有界约束下使得拉普拉斯能量最小化。这样做可通过形状感知和局部化的方式传播图柄的影响，即使是对具有复杂、凹边界的物体。变分权重优化还能够实现自定义权重，以使权重能够保持特定的重要物体特征的形状。我们证明了我们的混合权重可成功应用于 2D 和 3D 形状的实时形变。

1. 简介

交互式形变是指协助用户更改物体形状的任务。对于 2D 卡通形变，我们可以要求用户手动重新放置图像的每个像素，但是这非常繁琐，毫无必要。2D 形状的连贯配置空间要比图像每个像素所有可能的位置构成的空间小得多。因此，我们宁愿用户只提供少数几个高级约束，如“打开嘴巴”、“增大肚皮”或“弯曲尾巴”（图 1）。形状的其余部分应立即产生直观的形变。我们可借助图柄结构（如由刚性骨骼构成的骨架、封闭罩和和选定的区域或点）形成这样的高级约束与用户之间的接口。

有了这些图柄，交互式空间形变将成为编辑光栅图像、矢量图形、几何模型和动画人物的有效方法。如此众多的可能性产生了试图通过实时计算和直观使用改进交互式形变的大量方法。实时性能对于交互式设计（其中的任务需要创造性探索）和交互式动画（需要反复计算形变，通常每秒 60 次或更多）都至关重要。在所有形变方法中，线性混合及其变体因其速度较快而在实际使用中占据主导地位：物体上的每个点通过少量仿射变换的加权组合进行变换。

图 1. 我们的形变方法支持控制图柄（如点、骨骼或罩）的任意组合。这样的灵活性允许用户选择正确的工具：骨骼控制刚性部件，罩允许改变区域形状，点变换柔软部件。每个图柄的影响力权重在绑定时预先计算，因此可以很低的 CPU 使用率实时计算高品质形变。在本文中，着色标架展示在点图柄处指定的线性变换。

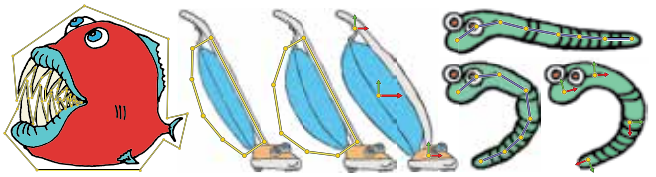


在典型工作流中，用户构造多个图柄，而形变系统将物体绑定到这些图柄；术语称为*绑定时 (bind time)*。然后用户操作图柄（交互式或通过编程方式），而系统相应地改变形状；这称为*定姿时 (pose time)*。遗憾的是，现有线性混合方案并不总是容易使用。用户必须*先验地*选择特定的图柄类型，不同的类型有不同的优势（图 2）。基于骨架的形变提供了对刚性肢体的自然控制，但对柔软的部位不太方便。广义重心坐标较之经典的基于格点的自由形式形变改进了对体积或面积的控制，但是仍需要构造完全封装形变物体的封闭罩或近封闭罩，处理起来非常繁琐。相比之下，变分技术通常支持点或区域上的任意图柄，但是定姿时开销更高。

如果能支持上述所有图柄类型（即点、骨架和罩），实时物体形变将更加容易。点可快速放置，且易于操控。它们指定平滑传播到物体邻近区域的局部形变属性（位置、旋转和缩放）。骨骼使某些方向比其他方向刚性

本文的原始版本发表于 *Proceedings of SIGGRAPH*，2011 年 7 月 11 日，ACM。

图 2. 不同的用户控制结构适合某些情况，但是不适合其他情况。设置封闭罩可能既繁琐又不直观：围绕*水虎鱼*的牙齿来回穿梭。罩是*吸尘器*的精确接合所必需的，这种情况下，在点处的缩放过于粗糙。点提供对*蠕虫*的松散平滑控制，而骨架形变过于僵硬，且过度复杂。



更强。如果两点间的部位太过易于弯曲，骨骼可将其变换为刚性肢体。罩允许一次影响物体的相当一部分区域，使得控制相关部位的膨胀和瘦化更加容易。

我们的目标是为线性混合方案提供影响力权重，从而对任意拓扑的图柄产生平滑直观的形变（图 1）。我们需要的是支持高解析度图像和网格 (mesh) 的实时交互式形变。我们需要点和其他图柄附近的平滑形变，这样它们就可直接放置在动画表面和扭曲纹理之上。我们为每个图柄寻求一个局部支持区域，确保每个图柄的影响能支配其邻近区域，并在其他图柄控制的物体区域中消失。

我们的解决方案通过最小化受上下界约束的拉普拉斯能量来自动计算混合权重。由于相关的欧拉 - 拉格朗日方程是双调和的，因此我们将这些权重称为*有界双调和权重*，而将得到的形变称为*有界双调和混合*。这些权重仅在绑定时计算一次。而在定姿时，物体上的点通过混合少量仿射变换来进行实时形变。我们的示例表明，有界双调和混合可产生平滑形变，并且点、骨骼和罩都有直观的局部影响，即使在具有复杂、凹进边界的物体上依然如此。我们的权重计算需要空间离散化和优化，这在某些应用中可能是缺陷，但我们公式的通用性可提供对能量最小化的额外控制，例如，定义保形权重以保持特定的重要物体特征的形状。

2. 以前的研究

众所周知，变分（或称能量最小化）方法可计算表面上任意图柄的高质量保形形变，[4, 6, 9, 22]，而某些变分方法对骨骼 [25] 起作用，或者可扩展到其他表面外图柄。[5] 这些技术的主要缺陷是它们依赖于定姿时优化。虽然系统矩阵可预先分解，并且回代可在 GPU 上实现，但这并不是类似于线性混合蒙皮之类的极易并行问题，因此慢得多。即使有显著的性能调优 [19] 或模型精简，[7, 23] 定姿时优化仍然过慢，无法以高帧速（例如视频游戏所需要的帧速）进行高解析度物体的形变。

大多数在定姿时十分快速的方法都使用图柄变换的加权混合来计算物体上每个点的变换。为了执行混合，有些方法使用移动最小二乘法，[16] 有些使用对偶四元数，[14] 但大多数方法使用线性混合蒙皮 (LBS)。[15] 使用 LBS 时，将图柄的仿射变换以不同的权重进行线性平均来变换每个顶点。虽然对旋转进行线性混合会导致众所周知的失真 (artifacts)，但 LBS 已在长达 20 多年的时间里成为骨架动画的流行技术，因为它简单、可预测，并且可在 GPU 上实现非常高效的定姿时计算。不仅骨架动画，大多数基于罩的形变方法 [8, 12, 13] 实际上也是 LBS，其中图柄（罩的顶点）变换被局限为平移，而重点是权重的选择。此外，上面提到的精简模型变分形变方法使用 LBS 来从精简模型转变为完整模型。

LBS 的权重选择决定了图柄的仿射变换是否直观地影响形状。某些情况下会使用图柄结构的闭合形式权重，[13, 17] 但更多时候，这些权重是在绑定时预先计算的，或者由手工绘制。在 3.1 节，我们列出了 LBS 权重理想属性的公式，在 3.2 节，我们将在这些属性的背景下，讨论以前的权重选择方案。

3. 有界双调和权重

我们的目标是在任意图柄上混合仿射变换来定义 2D 或 3D 形状的平滑形变。假设 $\Omega \subset \mathbb{R}^2$ 或 $\mathbb{R}^3$ 表示给定形状 $\mathcal{S}$ 和罩控制结构（如果存在）的并集所包含的体积域。我们用 $H_j \subset \Omega, j = 1, \ldots, m$ 来表示（不相交）控制图柄。图柄可以是单个点、区域、骨架中的骨骼（H_j 包含骨骼线段上的所有点）或者罩上的顶点。用户为每个图柄 H_j 定义仿射变换 T_j，并且所有点 $\mathbf{p} \in \Omega$ 由其加权组合加以变形：

$$\mathbf{p}' = \sum_{j=1}^{m} w_j(\mathbf{p})\, T_j \mathbf{p}, \tag{1}$$

其中 $w_j: \Omega \to \mathbb{R}$ 是与图柄 H_j 关联的权重函数。请注意，罩通常理解为 2D 中的闭合多边形或 3D 中的多面体，它们包含 $\mathcal{S}$ 或其一部分，但我们的框架与罩拓扑无关，并将罩简单地视为一组单形的集合，仅要求这些单形在罩顶点平移时的变换是线性的。因此，罩可以是开放的（图 3）。我们不将罩面（2D 中的线段或 3D 中的三角形）视作图柄；它们接收线性权重，我们将在 3.1 节中看到。另外请注意，对于由关节连接的骨架骨骼，我们正式在共用关节的每一根骨骼上引入关节点（我们假设骨架从不分开，即共用一个关节的所有骨骼将

关节变换到同一位置）。实际上，我们将共用点上的权重限制为在重叠骨骼之间均匀分布，以最大化权重的对称度。

3.1. 公式化

我们提出将权重 w_j 定义为高阶形状感知平滑度泛函（即拉普拉斯能量）的极小化变量，并服从图柄插值和多个其他理想属性的约束：

$$\arg\min_{w_j,\ j=1,\ldots,m} \sum_{j=1}^{m} \frac{1}{2} \int_{\Omega} (\Delta w_j)^2 dV \tag{2}$$

服从：$w_j|_{Hk} = \delta_{jk}$ (3)

$w_j|_F$ 为线性 $\forall F \in \mathcal{F}_c$ (4)

$$\sum_{j=1}^{m} w_j(\mathbf{p}) = 1 \qquad \forall \mathbf{p} \in \Omega \tag{5}$$

$$0 \le w_j(\mathbf{p}) \le 1,\ j = 1, \ldots, m, \quad \forall \mathbf{p} \in \Omega, \tag{6}$$

其中 $\mathcal{F}_c$ 是所有罩面的集，δ_{jk} 是克罗内克函数。图 4 演示了为点图柄计算的 w_j 示例。

我们的权重函数 w_j 拥有以下属性，可实现直观的高品质变形。

平滑度：图柄处缺乏平滑度会在 2D 纹理形状中造成明显可见的失真现象（图 5）并妨碍在 3D 形状上直接放置图柄。请注意，通过变分法，最小化拉普拉斯能量 (2) 就相当于解欧拉 - 拉格朗日方程，本例中，即双调和 PDE：$\Delta^2 w_j = 0$。与此相当，我们可以将我们的混合权重公式化为线性化薄板能量的极小化变量，因为这会产生同样的双调和 PDE（请参阅 Botsch 和 Sorkine 等人的文章 [6]）。如果提出的边界条件平滑，则有界双调和权重在图柄处为 C^1 连续，而在其他地方为 C^∞ 连续。这种情况始终成立，但骨架关节和罩顶点除外：对于由关节连接的骨骼，权重在关节处不连续，因为 w_j 在骨骼 H_j 上是 1，而它在相邻的骨骼上必须是 0。但是，这不会导致实际形变的平滑度问题，因为关节总是由所有相连的骨骼变换到同一位置。

要实现期望的行为，需要罩面上的显式线性内插约束 (4)，否则当平移罩顶点时，罩面不会线性变形。罩面上的这些线性约束妨碍了罩顶点上权重的平滑度。因此，我们的形变在罩顶点处不平滑，但是在其他各处是平滑的，包括跨罩面。

非负性：负权重会导致反直觉的图柄影响，因为形状的负权重区域朝指定变换的反向移动。我们在 (6) 中显式强制非负性，否则，双调和函数（如 Botsch 和 Kobbelt[3] 中所述）将经常为负，即使所有边界条件都非负（右）。

形状感知度：通俗地讲，形状感知度意味着图柄与域 Ω 之间的直观相关度。图柄的影响应符合形状的

图 3. 斜塔形变（原图在左侧显示）。相比其他类型的图柄，罩提供了对区域更精确的控制。

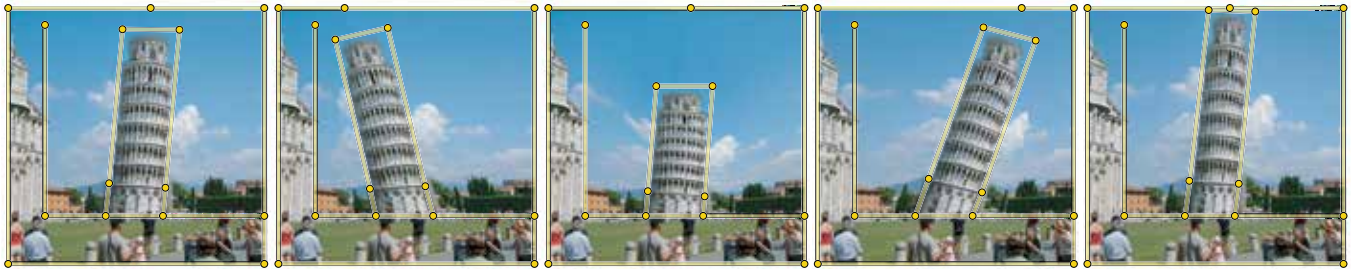
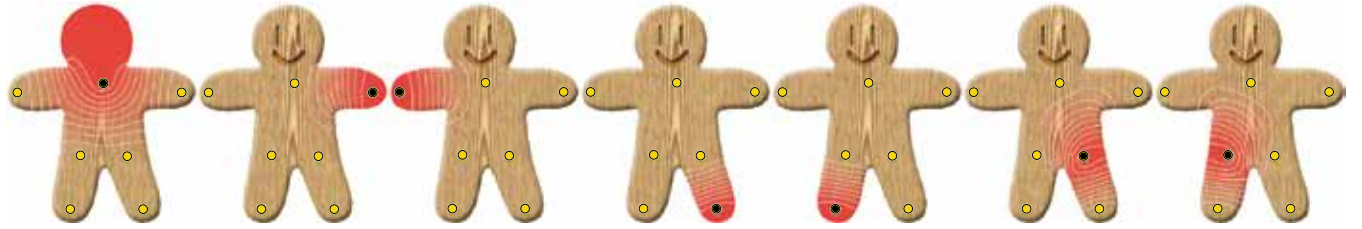
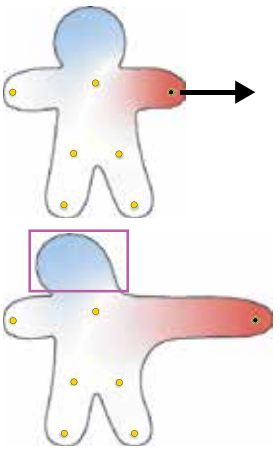


图 4. 有界双调和权重平滑且局部：每个图柄的混合权重强度以带白色等值线的红色部分显示。每个图柄对其接邻区域的影响最大，其影响在物体的遥远部分消失。



特征，并随着测地距离（而不是欧式距离）的增大而下降。最佳的形状感知行为是：权重 w_j 仅取决于 Ω 的度量，且不因任何可能的 Ω 嵌入而更改。我们的权重是形状感知的，因为双拉普拉斯算子只由度量确定。

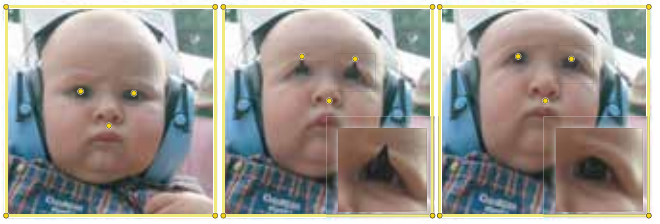


单位分解：此经典属性（也可在 Bézier 或 NURBS 中见到）确保如果同样的变换 T 应用到所有图柄，则整个物体将由 T 变换。我们在 (5) 中显式强制此属性，因为非负双调和权重总和不为 1，这跟无约束双调和权重不一样。

局部性和稀疏性：每个图柄应主要控制其邻近位置的形状特征，Ω 中的每个点应只受几个最接近的图柄的影响。具体来说，如果从点 $\mathbf{p}$ 到 H_j 的每一条局部最短路径（从形状感知的意义上说）都会通过某个其它图柄（附近），则 H_j “受阻”于 $\mathbf{p}$ 且 $w_j(\mathbf{p})$ 应为零。在我们的所有实验中，我们观察到了我们权重的这一属性。

无局部极大值：每个 w_j 在 H_j 上应达到其全局最大值（值为 1），并且不应该有其他局部极大值。此属性提供了图柄影响的单调衰减，并保证没有意外影响从图柄发生。此属性是在我们的所有测试中通过实验方式观察到的；通过施加边界约束 (6) 可能有利于此属性。若没有这些约束，双调和函数通常不一定在图柄处达到极大值，并会造成形变失真。虽然有界双调和权重*通常*没有局部极值，但它们肯定也并非总是单调的。[a] 实际上，局部极值总是出现在具有离图柄等（测地）距的长附属肢体的形状上。处理此非单调行为需要专门的优化。[11]

图 5. 权重必须处处平滑，尤其是在内部图柄处（很可能对应于重要特征）。在图柄处不连续的权重（如调和坐标（中）引起了撕裂失真（即使是对图柄轻微的变形）。我们的权重是平滑的，如右所示。



3.2. 与现有方案的比较

现有方案公式化并满足这些属性中的一部分，但非全部。例如，Shepard [17] 权重及类似权重（在嵌入式形变 [23] 和移动最小二乘法图像形变 [16] 中使用）是密集的，并且不是形状感知的。基于罩的方案不支持任意图柄：例如，将调和坐标 [12] 扩展到罩内部的图柄会导致缺乏平滑度（请参阅图 5）。热扩散权重 [2] 面临同样的问题。自然相邻插值 [21] 是少数几个保证局部性的方案之一，但是它在图柄处也不平滑。无约束双调和权重 [3] 是平滑的，但可能为负（或大于 1），有偏离图柄的局部极大值，并可能导致非局部影响。表 1 显示了多种方法满足的属性。

很多方法最近关注局部保留或规定角度。[24] 虽然它们在复分析方面有优雅的公式，但这些方法通常仅限于 2D。

3.3. 形状保持

能量最小化框架支持吸纳额外的能量项和约束来自定义权重函数。例如，一个有益的补充是，使指定区域 $\Pi \subset \Omega$ 的所有点经受同样的转换，即，让所有权重函数在 Π 上恒定 ($\nabla w_j|_\Pi = 0$)。因为我们通常只规定了图柄处的平移、旋转和统一缩放，因此这意味着 Π 将在 2D 中经历相似性变换，在 3D 中经历仿射变换，以使 Π 的形状得到保持。与 Igarashi 等人 [9] 的刚性刷相似，用户可以用刷（可能是软刷）绘制 Π，创建遮罩 $\rho : \Pi \to \mathbb{R}^+$；然后我们添加一个最小二乘项到能量中：

$$\sum_{j=1}^{m} \frac{1}{2} \int_\Pi \rho \|\nabla w_j\|^2 \, dV. \tag{7}$$

请见图 6，其中保形刷在变形鼻子时，帮助保持人

表 1. 选择混合权重的六种方法属性总结。我们的方法通常满足所有必要的属性。我们有经验证据，但是没有局部性和稀疏性的正式证明；我们的权重通常没有局部极大值。

属性	我们的	方法			
		[2,12]	[3]	[17]	[21]
平滑度	有	–	有	有	–
非负性	有	有	–	有	有
形状感知	有	有	有	–	–
单位分解	有	有	有	有	有
局部性，稀疏性	有*	–	–	–	有
无局部极大值	~	有，–	–	–	有

有 *：经验确认，~：经常，但非总是。

[a] 与我们最初发表的观察结果相反。

眼的形状。请注意，这跟放置图柄不同，因为没有显式变换需要由用户指定；绘制的区域根据其权重变换。

3.4. 实现

为了用二次规划数值求解，我们使用线性有限元离散化受约束的变分问题 (2)（我们对四阶问题使用扁平混合有限元方法 (FEM)，如 Jacobson 等人[10]的讨论）。假设物体 $\mathcal{S}$ 是 2D 多线段或 3D 三角形网格，我们在所有骨架的骨骼和罩面上采样得到顶点，并以与 $\mathcal{S}$ 中的所有图柄和顶点相容的方式对域 Ω 进行网格化。其结果是一个三角形 / 四面体网格 $\mathcal{M}$，其顶点 $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 包括所有离散化的 H_j 以及对象本身。权重变成了分段线性函数（我们正在寻找其顶点值）；我们用列向量 $\mathbf{w}_j = (w_{1,j}, w_{2,j}, \ldots, w_{n,j})^T$ 表示它们。

拉普拉斯能量 (2) 使用标准线性 FEM 拉普拉斯算子 $M^{-1}L$ 离散化，其中 M 是集中质量矩阵（每个对角顶 i 上有顶点 $\mathbf{v}_i$ 的 Voronoi 面积 / 体积 M_i），L 是对称刚度矩阵（即余切拉普拉斯算子）：

$$\sum_{j=1}^{m}\frac{1}{2}\int_{\Omega}(\Delta w_j)^2\,dV \approx \sum_{j=1}^{m}\frac{1}{2}(M^{-1}L\mathbf{w}_j)^T M(M^{-1}L\mathbf{w}_j) \quad (8)$$
$$= \frac{1}{2}\sum_{j=1}^{m}\mathbf{w}_j^T(LM^{-1}L)\mathbf{w}_j.$$

我们使用离散化图柄施加约束 (3)–(6)。为了离散化附加的保形能量项 (7)，我们采用线性 FEM 梯度算子 G（请见 Botsch 和 Sorkine[6] 的推导）。$G\mathbf{w}_j$ 是堆叠了梯度的向量，每个元素（2D 中的三角形和 3D 中的四面体）一个梯度；因为我们处理线性元素，所以每个元素上的梯度是一个常量）。设 R 是包含每个元素上用户刷 ρ 的积分的对角矩阵，设 $\bar{M}$ 是每元素质量矩阵（即对于每个三角形 / 四面体 i，$\bar{M}_i$ 包含其面积 / 体积）。然后 (7) 中的能量项被离散化为

$$\sum_{j=1}^{m}\frac{1}{2}\int_{\Omega}\rho\|\nabla w_j\|^2\,dV \approx \sum_{j=1}^{m}\frac{1}{2}\mathbf{w}_j^T(G^T R\bar{M}G)\,\mathbf{w}_j. \quad (9)$$

请注意矩阵 $G^T R\bar{M}G$ 是某种加权线性 FEM 拉普拉斯算子，因此其稀疏模式是主能量矩阵 $LM^{-1}L$ 的子集，不产生任何新的非零值。因此，添加此能量项不会增加优化复杂性。

我们使用 Triangle[18] 进行 2D 约束的 Delaunay 网格化，并使用 TetGen[20] 进行带约束的四面体网格化，以此创建离散化域。在 2D 中，我们配置 Triangle 来创建大小和形状接近统一的三角形。对于我们的所有 2D 示例，Triangle 只用了不到一秒，即使是需要像素大小三角形的细节图像也是如此。在 3D 中，我们配置 TetGen 来创建有粒度等级的四面体网格以降低复杂性（图 7）；对于具有 43,234 个顶点的*犰狳怪*网格以及沿骨骼内部采样的 120 个顶点，得到的四面体网格有 46,898 个顶点。对于*犰狳怪*以及我们的所有 3D 示例，TetGen 只用了几秒。

图 6. 我们优化框架的通用性能够计算保留显著物体特征的权重。标记一个区域以保留眼睛形状（中），否则眼部会扭曲（右）。

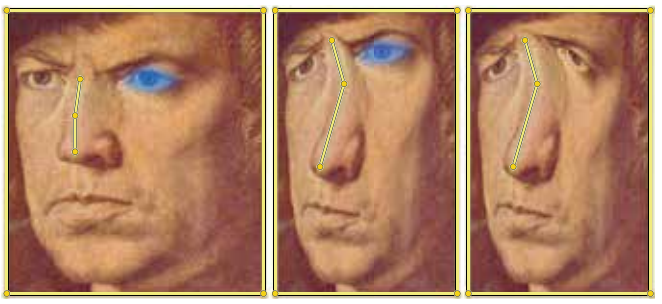


图 7. 嵌入*犰狳怪*的人体骨架不控制尾巴或耳朵，但是可以很容易地添加点来提高每个姿势的表现力。左：*犰狳怪*的一个剖面显示了 TetGen 生成的有粒度等级的四面体网格。内部四面体比表面附近的四面体大得多。这使得离散化的复杂度保持在合理范围内。

(8) 和 (9) 中的能量项相对于未知数 $\mathbf{w}_j$ 是二次的且是凸函数，而 (3)–(6) 是线性等式和不等式约束。我们使用 MOSEK[1] 作为稀疏二次规划求解器来同时计算所有图柄的权重。

由于解二次规划的必要时间与未知数数量呈超线性关系，因此将其分成多个更小的子问题可以显著提升速度。请注意，如果我们丢弃单位分解约束 (5)，则每个图柄的优化与其余图柄无关。我们实现了此策略，单独对每个 $\mathbf{w}_j$ 求解，然后在后处理时规一化每个顶点的权重。我们观察到此快速解与原始解之间的平均差几乎可忽略，这通常产生视觉上无法区分的形变效果（图 8）。更大的差偶尔发生在远离图柄的地方，但是权重有相同的定性行为：平滑度以及观察到的局部支持。例如，对于图 9 的*滴水兽*，分别计算 7 个图柄的权重比同时计算它们快 50 倍。我们将在第 4 节报告计时以及原始权重与这些快速权重之间的差。

算出权重后，形变本身是实时的（即使是很大的网格），因为这是用线性混合蒙皮的 GPU 实现计算的 (1)。

在基于罩的体系内各种重心坐标方法中，唯一的输入是罩顶点的平移。在我们的系统中，用户在每个图柄上提供完全仿射变换。根据具体应用的不同，用户可以选择只指定平移，即单位旋转和缩放。但是，非平凡旋转通常是实现满意的效果所必需的，而手动指定这些旋转可能非常繁琐。我们发现在采用有界双调和权重的线性混合蒙皮的精简子空间中，使用运行时优化从用户提供的平移推断出旋转会更加容易，这在精神上与 Der 等人 [7] 类似。

图 8. 放弃单位分解约束 (5) 将大大优化我们权重的预计算，而不失品质。左：每个顶点处原始权重与快速权重之间的平均绝对差，基于该顶点处的所有图柄权重。使用我们的原始权重（中）和快速权重（右）的相同图柄配置的形变视觉上无法分辨。

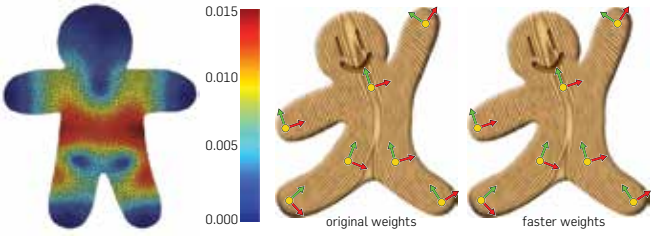


图 9. 滴水兽使用内部骨架和点图柄变形。

4. 结果

有界双调和混合将直观交互与实时性能结合在一起。其控制统一了三种不同的交互隐喻，使得简单任务仍旧简单，复杂任务更容易实现。

实验。点是用于操作柔软物体的特别精妙的隐喻。[9] 虽然可以用骨骼实现类似的变换，但图 10 中的*章鱼*和图 2 中的*蠕虫*证明了对易弯曲部位进行直接点操作的简便性，并表明对于同样的任务不适合使用刚性骨骼。

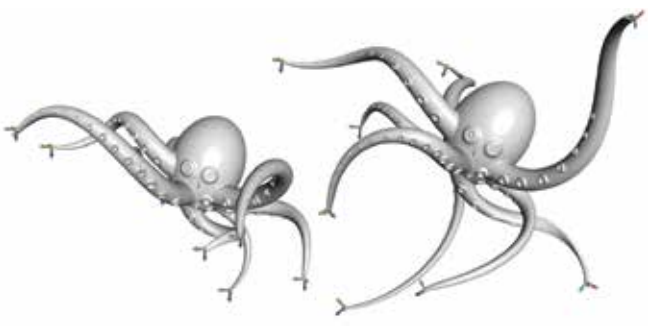
与以前的技术相比，[9,12] 我们的方法即使在图柄变换很大时，仍可平滑地形变。图 5 演示了平滑度对最小化纹理撕裂的重要性。

我们观察到我们的权重是局部的，在测试的所有示例中没有伪局部极大值。图 11 将我们的权重的支持区域与 Botsch 和 Kobbelt[3] 的双调和函数作比较，后者是全局支持的，并包含很多局部极值。

通过同时控制点和线，某些任务更容易完成。图 12 演示了在外部罩保持或调整图像边界时，我们的权重及平滑的点扭曲效果。罩非常适合精确面积控制。在图 3 中，我们使用一组任意的开放线和闭合线集合来操控塔的形状和方向。这些形变和精细调整涉及透视扭曲，很难只用点或只用线实现。

我们的方法可以很自然地推广到 3D。在绑定时，优化将权重遍布整个体，使得线性混合在运行时能产生平滑形变。此方案确保实时性能和低 CPU 利用率，即使对高解析度网格也是如此。我们注意到在 3D 中

图 10. 点用来衔接柔软的章鱼。



设置罩要比在 2D 中更加繁琐，尤其是需要罩来完全包裹物体的时候。对于操控图 13（左）中所示的手之类的任务，骨架更容易嵌入，更容易用来操控 3D 物体。骨架仍存在关节崩溃问题，并缺乏罩提供的精确体积控制，而我们的方法支持并简化了骨骼和罩的结合使用。特别要指出的是，我们的方法支持控制物体部件的局部罩（但不需要完全围住它）。图 13（右）显示了用来增大老鼠腹部的简单罩。与往常一样，局部罩可与点和骨骼结合，如此结合使用所有三种隐喻往往效果最佳。

图 7 演示了点和骨架的结合使用。我们通过嵌入人类骨架创建了犰狳怪的一系列姿势。骨架不控制尾巴和耳朵，所以它们的形变需要调整。这主要通过附加几个点来直接完成。直接表面操控便于将尾巴弯成更逼真的姿势，并生动地弯曲耳朵。类似地，图 9 中点图柄是控制滴水兽翅膀的展开和弯曲的自然而简单的选择。有界双调和权重将骨架的运动与这些点的配置相结合来产生平滑的形变。

讨论。我们在装有 8 GB 内存的 MacPro 四核 Intel Xeon 2.66 GHz 计算机上测试了我们的方法。表 2 报告了未优化代码的绑定时测量值。我们解决方案的一个局限是绑定时计算权重所需要的优化时间。我们使用线性 FEM 来离散化问题，虽然其他选择可能更高效，例如 Botsch 等人 [5] 和 Joshi 等人 [12] 使用的多解析度框架。在 3D 中生成有界双调和权重需要进行体离散化。请注意，一旦体积被计算之后，任意嵌入的物体（例如多边形集合 (polygon soup)）可随意变形，而与其拓扑无关。

我们的有界双调和权重没有线性精度属性，即，它们未必重现线性函数。此属性是仅通过内插罩顶点的位置来应用形变的基于罩的形变方法（如 Joshi 等人 [12] 和 Ju 等人 [13] 所述）所必需的，否则当罩旋转时，它们会扭曲形状。相比之下，我们的方法允许在图柄处提供任意变换，并在整个形状上混合它们；因此我们无需依赖线性精度也能处理旋转。线性精度调和坐

图 11.50 个点图柄（黑色和黄色）随机放在方形域中。左：黑色图柄无约束双调和权重的正负性（红色表示正区域，蓝色表示负区域）。局部极大值和极小值分别显示为红点和蓝点。左: 黑色图柄无约束双调和权重的正负性（红色表示正区域，蓝色表示负区域）。局部极大值和极小值分别显示为红点和蓝点。右：黑色图柄有界双调和权重的支持区域。在此测试及所有其他测试中，权重是局部的。这里没有伪局部极大值。

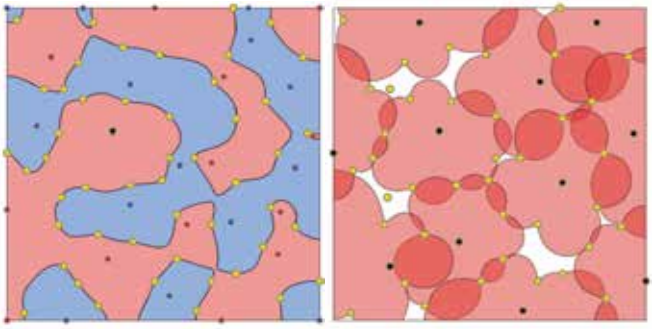


图 12. 点图柄通过混合在每个图柄处指定的仿射变换来使图像变形。边界上的罩保持矩形图像形状或允许其调整大小。

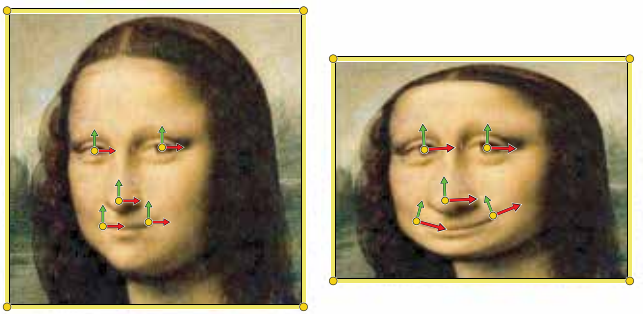


图 13. 完全包住 3D 物体（如手）的罩很难设置。骨架通常更容易嵌入和操作。当精确体积控制需要罩时，我们的方案允许只覆盖物体一部分的局部罩，从而使其更容易使用。

表 2. 本文各个示例的统计数据。|S| 是 3D 输入模型的三角形数量，|Ω| 是 Ω 离散化中的元素数量，*BT/h* 是每个图柄的绑定时间，单位为秒。E_{mean} 和 E_{max} 分别是我们显式强制执行 (5) 的原始权重与我们的快速权重（每个图柄的权重单独求解，然后规一化）之间的平均和最大绝对差。平均值和最大值是同时根据图柄和顶点计算的。

	|S|	|Ω|	BT/*h*	E_{mean}	E_{max}
姜饼人		5,040	0.1397	0.0043	0.058
邹眉		5,442	0.0906	0.0045	0.090
鳄鱼		7,019	0.1779	0.0013	0.055
比萨斜塔		12,422	0.3174	0.0025	0.060
蒙娜丽萨		32,258	1.2417	0.0050	0.11
滴水兽	20,000	46,003	1.1939	0.0043	0.18
手	28,692	51,263	3.1268	0.0020	0.37
老鼠	26,294	112,355	8.4464	0.0041	0.11
犰狳怪	86,442	142,073	12.0870	0.0041	0.40

标[12]与我们的罩之间的平移形变比较在图 14 用矩形图像显示。

本文中，我们只基于 (1) 实验了线性混合形变；但是，我们的权重对更高级的变换混合方法也很有用，如对偶四元数。[11, 14]

5. 总结

我们演示了如何统一所有常见类型的控制图柄，实现基于实时混合的形变的直观设计。这将允许用户为每项任务自由选择最便利的图柄，并使用户从手动绘制混合权重的繁重任务中解脱出来。

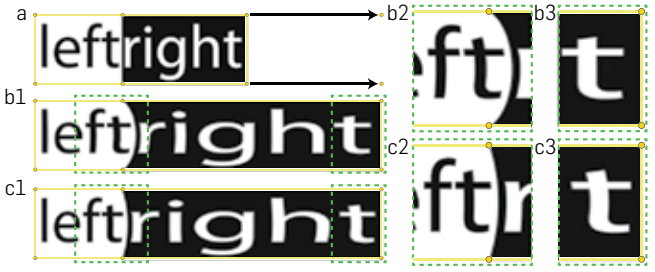
在未来的研究中，我们希望优化我们绑定时预计算的效率。除了寻求替代的离散化和数值方法外，进一步分析将有助于减少二次规划的维数：观察到的局部属性意味着权重在域的大量区域消失，因此可以从极小化中去除。

基于蒙皮的形变很容易折叠和自交，因为形变映射并不总是单射的。我们的方法也不例外。在此背景下，用我们的权重构建精简模型来进行模拟和接触处理值得探讨。我们还打算研究有界双调和权重的数学属性，以确定观察到的局部性和极大值原理成立的必要条件。

鸣谢

我们感谢 Jaakko Lehtinen、Bob Sumner 和 Denis Zorin 很有启发的讨论，感谢 Scott Schaefer 的*姜饼人*和*比萨斜塔*图片，感谢 Yang Song 帮助实现刚性刷和 2D 网格重划分。本文部分得到 NSF 奖 IIS-0905502、ERC 奖助金 iModel (StG-2012–306877)、SNF 奖 200021_137879、Intel Doctoral Fellowship 以及 Adobe Systems 馈赠的支持。 C

图 14. 我们显示了在罩内局部性与线性精度之间的权衡。文本图像 (a) 的静止姿势使用调和坐标 (b1) 和我们的有界双调和权重 (c1) 进行水平拉伸。调和坐标的响应比我们的响应 (c2) 更全局化 (b2)。另一方面，调和坐标保持了变形图柄旁边字母 T 中的竖线 (b3)，而我们的权重显示其缺乏线性精度 (c3)。



参考资料

1. Andersen, E.D., Andersen, K.D. The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. *High Performance Optimization*. H. Frenk, C. Roos,T.Terlaky, and S. Zhang, eds.Kluwer Academic Publishers, 2000, 197–232.
2. Baran, I., Popović, J. Automatic rigging and animation of 3D characters. *ACM Trans.Graph.26*, 3 (2007), 72:1–72:8.
3. Botsch, M., Kobbelt, L. An intuitive framework for real-time freeform modeling. *ACM Trans.Graph.23*, 3 (2004), 630–634.
4. Botsch, M., Pauly, M., Gross, M., Kobbelt., L. PriMo: coupled prisms for intuitive surface modeling.In *Proceedings of SGP* (2006), 11–20.
5. Botsch, M., Pauly, M., Wicke, M., Gross, M. Adaptive space deformations based on rigid cells. *Comput.Graph. Forum 26*, 3 (2007), 339–347.
6. Botsch, M., Sorkine, O. On linear variational surface deformation methods. *IEEE TVCG 14*, 1 (2008), 213–230.
7. Der, K.G., Sumner, R.W., Popović, J. Inverse kinematics for reduced deformable models. *ACM Trans. Graph.25*, 3 (2006), 1174–1179.
8. Floater, M.S.Mean value coordinates. *Comput.Aided Geom.Design 20*, 1 (2003), 19–27.
9. Igarashi, T., Moscovich, T., Hughes, J.F. As-rigid-as-possible shape manipulation. *ACM Trans.Graph.24*, 3 (2005), 1134–1141.
10. Jacobson, A., Tosun, E., Sorkine, O., Zorin, D. Mixed finite elements for variational surface modeling.In *Proceedings of SGP* (2010).
11. Jacobson, A., Weinkauf, T., Sorkine, O. Smooth shape-aware functions with controlled extrema.In *Proceedings of SGP* (2012).
12. Joshi, P., Meyer, M., DeRose, T., Green, B., Sanocki, T. Harmonic coordinates for character articulation. *ACM Trans.Graph.26*, 3 (2007).
13. Ju, T., Schaefer, S., Warren, J. Mean value coordinates for closed triangular meshes. *ACM Trans.Graph.24*, 3 (2005), 561–566.
14. Kavan, L., Collins, S., Zara, J., O'Sullivan, C. Geometric skinning with approximate dual quaternion blending. *ACM Trans. Graph.27*, 4 (2008).
15. Magnenat-Thalmann, N., Laperrière, R., Thalmann, D. Joint-dependent local deformations for hand animation and object grasping.In *Graphics Interface* (1988), 26–33.
16. Schaefer, S., McPhail, T., Warren, J. Image deformation using moving least squares. *ACM Trans.Graph.25*, 3 (2006), 533–540.
17. Shepard, D. A two-dimensional interpolation function for irregularly-spaced data.In *Proceedings of ACM National Conference* (1968), 517–524.
18. Shewchuk, J.R.Triangle:Engineering a 2d quality mesh generator and delaunay triangulator.In *WACG* (1996), 203–222.
19. Shi, X., Zhou, K., Tong, Y., Desbrun, M., Bao, H., Guo, B. Mesh puppetry: cascading optimization of mesh deformation with inverse kinematics. *ACM Trans.Graph.26*, 3 (2007), 81:1–81:10.
20. Si, H. TETGEN: a 3D delaunay tetrahedral mesh generator, 2003. http://tetgen.org.
21. Sibson, R. A brief description of natural neighbor interpolation. *Interpolating Multivariate Data*. V. Barnett, ed. Volume 21, John Wiley & Sons, 1981, 21–36.
22. Sorkine, O., Alexa, M. As-rigid-as-possible surface modeling.In *Proceedings of SGP* (2007), 109–116.
23. Sumner, R.W., Schmid, J., Pauly, M. Embedded deformation for shape manipulation. *ACM Trans.Graph.26*, 3 (2007), 80:1–80:7.
24. Weber, O., Ben-Chen, M., Gotsman, C., Hormann, K. A complex view of barycentric mappings. *Comput. Graph.Forum 30*, 5 (2011).
25. Weber, O., Sorkine, O., Lipman, Y., Gotsman, C. Context-aware skeletal shape deformation. *Comput. Graph.Forum 26*, 3 (2007), 265–274.

Alec Jacobson 和 Olga Sorkine-Hornung {jacobson, Sorkine}@inf.ethz.ch)，瑞士苏黎世 ETH Zurich。

Ilya Baran (baran37@gmail.com)，马萨诸塞州波士顿 Belmont Technology Inc.。

Jovan Popović (jovan@adobe.com)，华盛顿州西雅图 Adobe Systems, Inc.。

译文责任编辑：周昆