

计算机 · 电子 · 通讯

# 技术评论

## 数据、机器与健康

- 智能设备定制个人医疗
- 医疗数据的保密技术
- 人体连上互联网
- 移动医疗：  
传统医院的自我改造
- 全基因组测序：  
革命疗法还是隐私噩梦？

合作机构



定价：25元

ISSN 1673-5153



# 云计算

人工智能

工控机

制造业

电子

嵌入计算



传感器

互联网

3D 打印

绿色计算

图形图像

虚拟现实



# 互联网金融

芯片

## 大数据

软件架构

纳米架构

# 人机交互

多媒体

## 普适计算 MEMS



微信名：计算人 微信号：jisuanren



# COPYRIGHT 版权

## 主管单位 Authorities in Charge

中华人民共和国教育部 Ministry of Education of the People's Republic of China

## 主办单位 Sponsor

中国大学出版社协会 China University Presses Association

## 出版单位 Publisher

《环球科学》杂志社有限公司 GLOBAL SCIENCE MAGAZINES Co., Ltd

社址 Address: 北京市朝阳区秀水街1号建外外交公寓4-1-21 Office 4-1-21, Jianguomen Diplomatic Residence Compound, No. 1, Xu Shui Street, Chaoyang District, Beijing, China. 邮编: 100600

联系电话: 010-85325810/85325871

## 社长 / 总编辑 Editor-in-chief

陈宗周 Chen Zongzhou

## 副社长 / 副总编辑 Deputy Editor-in-chief

刘芳 Liu Fang

## 指导机构 Adviser Agent

中国计算机学会

## 执行出版人 Publisher

管心宇 Xinyu Guan

## 资深编辑 Senior Editor

马法达 Falda Ma

刘妍 Yan Liu

## 特约编辑 Contributing Editor

史彦诚 Yancheng Shi

刘大明 Daming Liu

高天羽 Tianyu Gao

费錡 Yong Fei

王璇 Xuan Wang

## 运营中心 Operating Department

运营合作机构 Publisher

上海灵宸文化传媒有限公司

发行部 Circulation Department

发行总监 Circulation Director

谢磊 Xie Lei 010 - 57439192

市场部 Marketing Department

市场总监 Marketing Director

孔祥彬 Kong Xiangbin 010 - 85325810 - 807

广告部 Advertising Department

销售总监 Sales Director

范欢 Fan Huan 010-85325871-802 010-85325981

读者服务部 Reader Service

杜珺 Du Jun 010 - 57458982

印刷: 北京利丰雅高长城印刷有限公司

如发现本刊缺页、装订错误和损坏等质量问题, 请在当月与本刊读者服务部联系调换(请将坏书寄回)。

国际标准刊号: ISSN 1673-5153

国内统一刊号: CN11-5480/N

广告经营许可证号: 京朝工商广字第8144号

## 知识产权声明:

IEEE, IEEE Computer, IEEE中文网站的名称和标识, 属于位于美国纽约的电气电子工程师学会有限责任公司所有的商标, 仅通过授权使用, 这些材料的一部分由IEEE Computer英文版翻译而来, 版权归IEEE所有, 并经IEEE授权翻译复制。

IEEE Computer杂志的中文版权, 由美国电气电子工程师学会有限责任公司授予上海灵宸文化传媒有限公司, 并由本刊独家使用。

本刊发表的所有文章内容由作者负责, 并不代表上海灵宸文化传媒有限公司、美国电气电子工程师学会有限责任公司的立场。

本刊内容未经书面许可, 不得以任何形式转载或使用。

## 北京市绿色印刷工程 ——优秀青少年读物绿色印刷示范项目

## IEEE Computer Society

### 编辑部

执行编辑

Carrie Clark Walsh

ccwalsh@computer.org

编辑

Mark Gallaher

资深新闻编辑

Lee Garber

资深编辑

Chris Nelson

特约编辑

Christine Anthony

多媒体编辑

Brian Brannon

Ben Jones

Erica Hardison

设计和产品

Jennie Zhu-Mai, Lead

Monette Velasco

Alex Torres

设计

Olga D'Astoli

封面设计

David Angel

### 管理人员

产品和服务总监

Evan Butterfield

编辑服务

Robin Baldwin

资深商务开发经理

Sandy Brown

资深广告协调

Marian Anderson

### 主编

Ron Vetter

北卡罗来纳大学威尔明顿分校 vetterr@uncw.edu

### 副主编

Sumi Helal

佛罗里达大学 helal@cise.ufl.edu

### 副主席, 学术研究

Kathleen Swigger

北得克萨斯大学 kathy@cs.unt.edu

### 副主席, 特刊

Bill N. Schilit

谷歌公司 schilit@computer.org

### 计算应用编辑

Rohit Kapur

Synopsys rohit.kapur@synopsys.com

### 展望栏目编辑

Bob Colwell

bob.colwell@comcast.net

### 多媒体编辑

Charles R. Severance

csev@umich.edu

### 2014 年 IEEE 计算机协会主席

Dejan S. Milojicic

d.milojicic@computer.org

### 行业编辑

计算机架构

David H. Albonesi

康奈尔大学

Greg Byrd

北卡罗来纳州立大学

图形和多媒体

Oliver Bimber

奥地利约翰开普勒林茨大学

高性能计算

Vladimir Getov

韦斯特敏斯特大学

信息和数据管理

Naren Ramakrishnan

弗吉尼亚理工大学

互联网计算

Simon Shim

圣何塞州立大学

多媒体

Savitha Srinivasan

IBM 阿尔玛登研究中心

网络

Ahmed Helmy

佛罗里达大学

Ying-Dar Lin

台湾国立交通大学

安全和隐私

Rolf Oppliger

eSECURITY Technology 公司

软件

Renée Bryce

北得克萨斯大学

Jean-Marc Jézéquel

雷恩大学

David M. Weiss

衣阿华州立大学

### 专栏编辑

云封面

San Murugesan

BRITE Professional Resources

计算对话

Charles R. Severance

密歇根大学

分析

Naren Ramakrishnan

佛吉尼亚理工大学

教育

Ann E.K. Sobel

迈阿密大学

娱乐计算

Kelvin Sung

华盛顿大学博赛尔分校

绿色 IT

Kirk W. Cameron

佛吉尼亚理工大学

识别科学

Karl Ricanek

北卡罗来纳大学威尔明顿分校

研发

Chris Huntley

菲尔费尔德大学

无形计算

Albrecht Schmidt

斯图加特大学

带外 (Out of Band)

Hal Berghel

内华达大学拉斯维加斯分校

科幻小说原型

Brian David Johnson

英特尔

### 安全

Jeffrey M. Voas

国家标准技术研究所

社会计算

John Riedl

明尼苏达大学

软件技术

Mike Hinchey

Lero—爱尔兰软件工程研究中心

32/16 年前

Neville Holmes

塔斯马尼亚大学

### 顾问委员会

Carl K. Chang

荣誉主编

Iowa State University

衣阿华州立大学

Jean Bacon

剑桥大学

Hal Berghel

内华达大学拉斯维加斯分校

Doris L. Carver

路易斯安那州立大学

Naren Ramakrishnan

弗吉尼亚理工大学

顾问

Alf Weaver

弗吉尼亚大学



郑州国际会展中心 10月23~25日

**国内计算领域最大的学术盛会**  
**2014中国计算机大会**  
China National Computer Congress 2014

主办单位：中国计算机学会

In Cooperation with:



承办单位：信息工程大学 郑州市人民政府



Ivan Sutherland

ACM图灵奖获得者  
计算机图形学和虚拟现实之父



赵沁平

CCF会士 虚拟现实技术专家  
北京航空航天大学教授



方滨兴

CCF会士 信息安全专家  
北京邮电大学教授



邬江兴

中国程控电话交换机之父  
信息工程大学教授



沈晓卫

IBM中国研究院院长



张 潼

百度深度学习研究院（IDL）首席科学家  
美国罗格斯大学教授



Sridhar Iyengar

英特尔研究院（Intel Labs）副总裁

登录大会网站

<http://cncc.ccf.org.cn> 报名参会

**CCF会员享受特别优惠！**

联系方式：010-62600336 cncc@ccf.org.cn

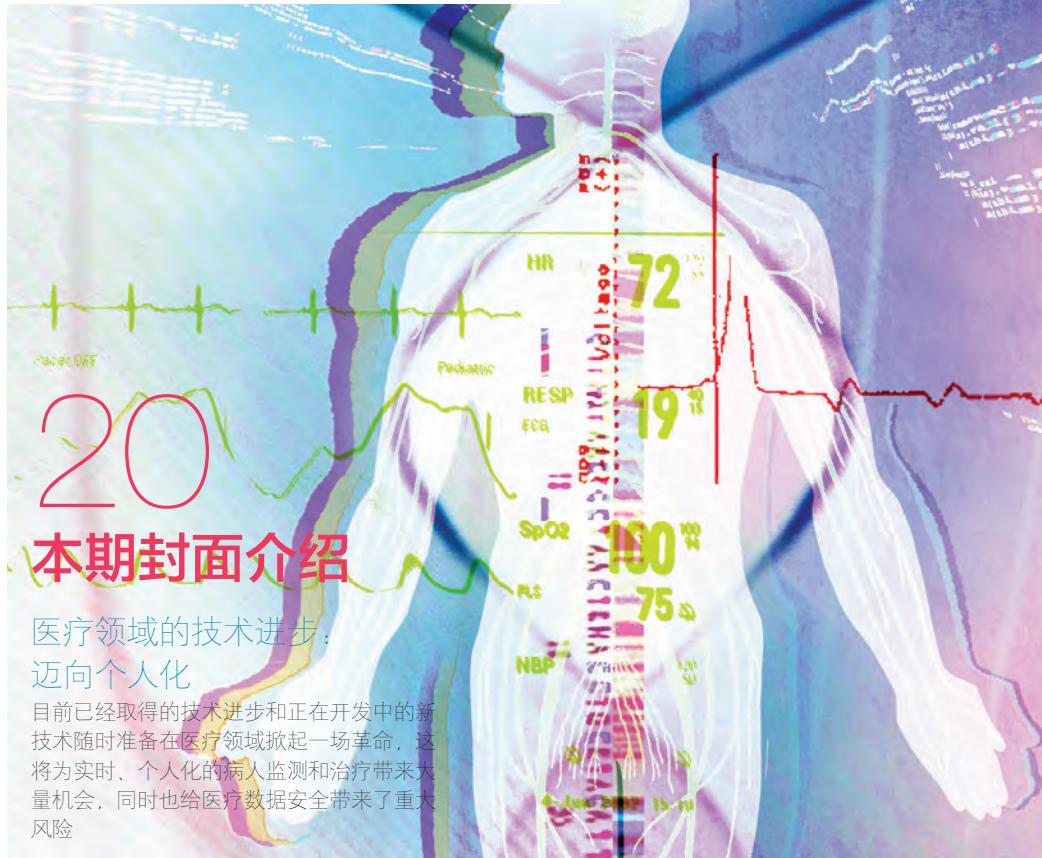


微信服务号  
中国计算机大会

立即申请  
加入CCF  
APRIL 2015



# 技术评论



2015年2月刊

## 封面报道

24

32

40

### 个人健康系统中的 隐私保护

个人健康记录是自行管理的医疗服务的组成部分，但记录的访问权由病人控制。这引起了人们的严重担忧，需要人们更好地平衡个人化、隐私保护和安全控制等要素。

### 实现疾病自我管理 的智能移动系统

廉价的移动医疗不仅要注重技术性能，而且要考虑病人和医疗供应商之间的责任划分，病人的病情背景，以及如何实现对病人决策的支持和定制化交流互动。

### 实现医用级别的实 时移动医疗服务

一项心电图无线监测的案例研究表明，当前的CDMA2000蜂窝通信技术在医疗遥测领域拥有可观潜力。相关网络协议栈经改良后，可实现最大的数据完整性和最短的服务延迟。

2015年2月刊

## 目录



50

58

### 医疗数据整合与云端信息学

用于医疗应用的信息整合和信息学框架，能够利用由商用硬件组成的云端大规模分布式批处理基础设施的并行计算能力，为高级医疗应用的开发者提供新的灵活性。

### 全基因组测序：医学革命还是隐私噩梦？

很快，全基因组测序就能让许多人都负担得起了。但棘手的隐私和伦理问题会妨碍它的推广，并阻挠基因组学在医疗领域的大规模应用，延缓可能出现的医学进步。

## 高端访谈

6 听听 NiSoft 的 CEO 在说什么

## 大师小传

10 编译器与 Cobol 之母：格蕾丝·霍普

## 新兴市场

14 哥伦比亚的人机互动教育

## 数据分析

70 分析学、机器学习和物联网

## 安全 IT

74 索尼的灾难会再来吗？

## 科幻小说原型

77 虚拟反乌托邦

## 卷末

80 加密到底应该是什么？



77

Editor: Joseph Williams, Seattle Pacific University, josephwi@spu.edu



# 与 NiSoft 首席执行官 道格·迪尔多夫 一席谈

约瑟夫·威廉姆斯 (Joseph Williams)，西雅图太平洋大学 (Seattle Pacific University)

**现**代企业的管理层中，许多 IT 专家的角色已经不局限于首席信息官 (CIO) 和首席技术官 (CTO) 了。首席信息官变身首席执行官的例子已不胜枚举：原任摩根大通 CIO 的盖伊·基亚雷洛 (Guy Chiarello) 出任第一资讯的 CEO；克里斯·洛夫格伦 (Chris Lofgren) 卸任施奈德物流的 CIO 之后经过一番辗转，最终成为了老东家的 CEO；前威讯的 CIO 沙伊甘·克拉德皮尔 (Shaygan Keradpir) 成了瞻博网络 (Jupiter Networks) 的 CEO。

还有的 IT 专家沿着其他道路走入了管理层：萨蒂亚·纳德拉 (Satya Nadella) 在升上微软 CEO 之前从事产品工程；谷歌 CEO 埃里克·施密特 (Eric Schmidt) 曾经是施乐和贝尔实验室的研发技术员；推特的全球营收总裁亚当·贝恩 (Adam Bain) 最初负责几家网络报纸的技术运营，后来一步步升到了现在的位置。

不过，还有一条晋升管理层的通道是较少有人提及的，那就是销售。要想

将 IT 解决方案成功推销出去就必须具备深思熟虑的头脑、要将商业和技术两方面的技艺和学问融为一体。做到了这一点，再加上勤恳工作、积累人脉，一位 IT 专家就能踏着销售的阶梯，一步步攀上管理层的高位——NiSoft 的 CEO 道格·迪尔多夫 (Doug Deardorf) 就是一个例子。

NiSoft 的业务是在全世界出售软件方案，从而保护大型工业设施的安全。眼下公司的软件正在全世界的 350 多处设施中发挥作用，其中既有世界最大的合成燃料设施（位于南非），又有中东第一家混合动力海水淡化厂。不久前，公司又签下了美国的一座电站，准备将它建设成全美第一家能够捕捉并且储存二氧化碳排放的发电设施。大体而言，NiSoft 的方案可以归为高端电子许可证管理系统，它们支撑着那些极其复杂的安全流程，从而为大型工业生产设施的运营保驾护航。这些方案类似于 IBM Maximo 或 SAP 的产品。因此，NiSoft 不仅有庞大的客户，也要面临强大的对手。

迪尔多夫于 2008 年出任 NiSoft 的

CEO。他在几年前曾为公司服务，后来得到机会，以主要投资者和 CEO 的身份回归。纵观他的职业生涯，一项主要的工作就是向基础设施和能源产业提供技术方案。他在国内外的工作中积累了广泛的经验，曾与艾创 (Itron)、普华永道、Indus 集团（现为 ABB 集团旗下的 Ventyx 公司）等企业合作。

迪尔多夫的 CEO 之旅始于在西雅图太平洋大学的求学时代，当时他一心钻研 PDP-1145，打算毕业后加入波音公司。但是在一次“实习面试”(practice interview) 中，一名来自宝来公司 (Burroughs) 的招聘员却告诉迪尔多夫比起购买螺丝和机身部件，还是卖电脑更有意思，也更赚钱。宝来为迪尔多夫提供了全面的行业技术和销售培训，为他在硬件、软件和系统集成方面的销售工作打下了扎实的基础。在宝来工作期间，迪尔多夫对国际事务越发关心，于是他暂别职场，在雷鸟国际管理学院 (Thunderbird School of Global Management) 攻读了 MBA 学位。在此之后，他又在几家技术

公司的国际部从事销售和管理工作，职位越来越高。

身为 NiSoft 的 CEO，迪尔多夫仍自诩为一个技术专家，然而因职责所限，他对公司的产品已经不可能深究其技术细节了。那么他是否怀念那些亲临一线的日子？是的，不过运营一家企业也有其需要亲手处理的难题，尤其是在他居住美国丹佛，而公司总部却远在北爱尔兰的情况下。

#### 初任 CEO 的时候，内心有什么感受？

兴奋和满足是首先涌现的情绪。我回到了几年前曾经服务的公司，如今的任务是主持收购并出任 CEO。多亏我的金融搭档新月资本（Crescent Capital，位于贝尔法斯特）的支持，降低了我接任 CEO 的难度。

NiSoft 服务的都是声誉卓著的企业，我从工作开始就常与这些企业往来。我将自己的金融地位和职业前途托付给 NiSoft，因为我信任那里的人，也看好公司的发展。

除此之外，我也深深体会到了 CEO 这份工作是何等的孤独。员工和客户的福利，最终都取决于企业 CEO 的抉择，其中的责任是巨大的。而且推动决策的信息也往往不能与人共享。

**身为 CEO，个人遇到的最大难题是什么？**  
给组织内部带来变化、又不能破坏现成的有效机制，这就是最大的难题。NiSoft 的公司的组成很有意思：一方面它规模很小，员工不到 50 人；但另一方面，它又在全球设有三个分部，并为超过 25 个国家的超大型公司提供复杂而不可或缺的软件应用。公司成立 22 年，一直是本领域一枝独秀的应用供应商。这固然是个非

## 道格·迪尔多夫：一点趣闻



**最近读过的好书：**丹尼尔·詹姆斯·布朗（Daniel James Brown）的《划船少年》（The Boys in the Boat），划过赛艇的人都该看看。书中记叙了华盛顿大学赛艇队 1936 年在国内外取得破天荒胜利的故事。

**最喜欢的行业刊物：**我订了石油天然气行业和电力行业的几个情报源，还长期订阅《金融时报》、《经济学人》、《外交》和《彭博商业周刊》。

**最近一次编写代码 or configured** 回答这个问题会暴露家庭隐私的！（好吧，是 1983 年。）

**如果不是一家软件公司的 CEO，你还会留在软件行业吗？**多半会的。要在企业软件行业成为能干的供应商或合作伙伴，你就要有过人的创意和自律。我觉得这很吸引我。

**你携带的是什么移动电话？** iPhone 5。

**你还会使用什么别的个人技术？** iPad、Garmin Edge（骑车时用）和 Garmin GPS（远足时用）。

**大学是哪里上的？** 我本科上的是西雅图太平洋大学，获得了商业管理和心理学的双学士，后来又在雷鸟国际管理学院获得了国际管理的 MBA 学位。

同小可的成就，但是我们仍可继续向前，一方面为现有的客户提供更加关键的技术，另一方面也向别的行业扩张。

司的员工、伙伴和客户留下怎样的印象？我对这一点是相当在意的。我的工作是服务这些人群，我不能给他们造成困扰、使他们觉得我的作为在改变公司的传统。

**在一家总部设在海外的企业担任 CEO，是怎样一种感觉？**

这种感觉非常吸引人。我非常珍惜这个环境中不断涌现的人脉和学习机会。物流方面的挑战大致可以通过技术来应付，但我是个喜欢交流的人，我深信长远来看，还是面对面的互动最有成效。所以身为 CEO，我常常需要出差。眼下美国分部的收入要比英国分部稍高一些（后者服务欧洲、中东和非洲），但我们仍是一家不可动摇的英国公司，总部也始终设在北爱尔兰的贝尔法斯特。我是美国人，在美国居住并管理公司，这会给公

**在你的工作中，“创新”处于什么位置？你会推动创新吗？**

当初启动收购，我对 NiSoft 的期望是能与客户建立更加深厚的联系。这样做能够增加收入，扩大利润，也能使公司更加稳定。因此就需要在全公司上下推动创新，从公司文化开始，进而推广到产品，再到我们服务市场的风格。我们花费了许多精力反思公司的经营方式。我在达拉斯生活过好些年，用当地的土话来说，我们需要的是一点“德州式自负”（Texas swagger）。我可以谨慎地说一句：我对

# 高端访谈 LIFE IN THE C-SUITE

公司的创新起到了不小的推动作用，但是我也发现，创新的意愿、人力和才能其实早就存在了，只是需要激发而已。有了金融搭档新月基金的大力支持，我们得以采取更加强势的行动，而它们的效果也的确不错。

**杰里·宋飞有纽曼这个对手，福尔摩斯有莫利亚蒂教授这个宿敌，你的眼中钉又是谁呢？**

我还真想不出有什么眼中钉！当然了，从正面的意义上看，我们的系统经理可以算作一个。他在贝尔法斯特驻扎，加

我至今还很喜欢这项工作。打通人脉、解决问题、合作共事、引入方案，一方面为客户改善业务，一方面也增加自身的经验和收入、促进公司的成长，这些都是驱使我的动力。

**你是如何把握技术发展的大体趋势的？**

我的主要学习手段是观察和倾听别人的解释，其次才是阅读。我会和公司同事、和客户谈论正在应用的技术，了解他们在思考什么、又对什么感兴趣。

对于 NiSoft 的产品线，你是如何把

安静的那个声音才是值得倾听的。

**对于有抱负成为管理者的 IT 专家，你有什么建议？**

这个说来有点老套：要向着理想而行。不要给自己设限，也不要让别人限制自己。要扩大知识，不能仅限于工作所需。要明白如何激励他人。自己的工作之外，还要花些力气帮助别人——这是很好的训练。

**跟我说说你的一天一般如何度过？你的日程很有规律吗？**

不，我的日程并不算规律。我的活动丰富多样，这至少有一部分是因为公司的经营范围遍布全球。

**你每小时接收多少封电邮？**

平均 10 封。贝尔法斯特和新加坡分部常在夜间发来邮件，因此我每天早晨上班时，邮箱往往是满的。今早我就收了 68 封。

**午饭怎么吃？会休息吗？**

除非是跟客户或潜在的客户吃饭、或者约好了午餐会，我一般都在办公桌前吃。我常吃的东西有水果、蔬菜、坚果和酸奶，边吃边看东西。

我的确会在上班时休息，但不是午饭时间。我喜欢每周去几次健身房、或者慢跑几次。这些都是极好的休息，可以在白天的任何时候进行。

**晚上会在家里完成多少工作？**

我经常在晚饭后工作二到四个小时，长短取决于开放项目的紧迫程度（所谓开放项目就是合约正在洽谈的项目），以及是否需要打电话去亚洲。如果工作到了 10 点以后，我可能就会干脆工作到凌晨 2 点。

## 第一，对团队要激励、装备，然后信任。第二，勤劳是成功的不二法门。

入公司已经有些年头了，在这之前，他还曾为几家规模远超我们的公司服务过。他总是静静地听我说话，一副不动声色的表情，然后靠到座椅背上说一句“哦，你的意思其实是……”接着就开始有理有据地订正我的想法。我们常常针锋相对地辩论，但是完全不伤感情。

**除了制定最高和最低纲领，你身为 NiSoft 的 CEO，遇到过什么别的重大难题？**  
为了履行向更大的市场提供产品与服务的使命，就必须维持并传达一个强势的姿态，但同时又要考虑到公司的限度、不可冒进，因为在重要期限之内完成客户的嘱托也是公司的义务。

**说一件你需要授权他人、却至今不曾放手的工作。**

答：销售管理。我的本行是企业销售，

握其技术细节的？

身处小公司的好处之一是能够参与公司业务的方方面面。我会参加产品的战略决策、决定使用什么样的技术工具、以及如何在全世界有效部署我们的方案。我常常参加会议，无论是关于产品的，还是关于客户项目推进的。

**身为 CEO，哪项工作给了你最大的满足？**

最满足的莫过于扮演导师，并亲眼看着组织第一次成功的情景，当同事们克服恐惧和怀疑，终于做成某事的时候，那种满足是最强烈的。我最喜欢大家欢呼“我们成功了！”的那一刻。

**工作至今，最宝贵的经验是什么？**

最宝贵的有三条。第一，对团队要激励、装备，然后信任。第二，勤劳是成功的不二法门。第三，许多时候，房间里最

#### 周末呢？

我尽量不在周末工作。我在这一点上越做越好，过去两年，我的周末越来越空了。

#### 你外出度假吗？如果去的话，你会连网吗？

是的，我很喜欢外出。有时是短途旅行，花两三天到山上滑雪，也有的时候是真正的度假，时间超过一周。我的确会在度假的时候通过电邮和手机与办公室保持联系，这么做有两个目的：一是查看和整理往来的电邮，以便在假期结束时顺利开始工作。二是让我的高级团队知道我联系得上，如果有重大或紧急的事情可以找我商量。我完全不会因为度假时还连着网络就坏了兴致。

#### 你觉得目前的步调你能维持多久？

我喜欢现在的工作。现在精力充沛、环境也好，能有所贡献，我很感激。我天生就是闲不下来的人，很难想象放慢了步调会是什么样子。不过是用工作换取另一天。

#### 你多久见一次 CTO、CIO 和 CFO？

首席财务官和我一样人在丹佛，只要我在丹佛，我们天天都见。公司的产品经理（CTO）和系统经理（CIO）常驻贝尔法斯特，我去贝尔法斯特时会和他们相处很久。我大概每六周去贝尔法斯特待上一周。我常和他们面谈或开视频会议，有时候单独商量，有时候集体开会，往往一天就要谈上几次。

#### 你一天中的什么时候状态最佳？

我不太确定我有“状态最佳的时候”（因

那也有代价。除了熬夜，我也很喜欢起个大早，趁着没到高峰的时候赶到公司，然后看着日出，在开会之前把邮箱里的来信处理完毕，那种赶在时间表之前的感觉使我非常满足、精力充沛。

#### 你觉得自己的下一份工作可能是什么？

我希望利用自己积累的经验和人脉再创造一点什么。我出身软件行业，如果还有“下一份工作”，或许也是帮助另外一家软件公司扩展业务。

(本文内容来自 *IT Professional*) 

我完全不会因为度假时还连着网络就坏了兴致。

为那样就表示也有“状态不佳”的时候，而我并没有那个感觉）。我天生是个夜猫子，晚上 10 点以后会再来精神。我在夜里 10 点到凌晨 2 点之间效率很高，但

**约瑟夫·威廉姆斯**，西雅图太平洋大学商业与经济学院院长，联系方式：  
josephwi@spu.edu。



# 会议就在你的手中

IEEE计算机协会的会议发布服务（CPS）现在可以提供组织会议的移动应用了！让会议的日程、会议信息和论文列表在你的与会者手中的设备上显示。



会议的移动应用可在**安卓**设备、**iPhone**、**iPad**和**Kindle Fire**上运行。

欲知更多信息，请联系 [cps@computer.org](mailto:cps@computer.org)

EDITOR: George Strawn, NITRD, gostrawn@gmail.com



## 编译器与 Cobol 之母： 格蕾丝 · 霍普

乔治 · 施特劳 (George Strawn)，网络与信息技术研发项目 (NITRD)  
坎迪斯 · 施特劳 (Candace Strawn)

**格**蕾丝·穆雷·霍普将军 (1906-1992) 是第一位著名的女性计算机科学家 (可能要排在 19 世纪的艾达拉芙蕾丝伯爵夫人之后，她为查尔斯·巴贝奇未建造完成的蒸汽动力计算机编写了程序，那是另一段故事了)。这篇霍普的小传主要关注她早期的编程生涯，她如何构建了第一个编译器，领导创立 Cobol 语言，以及她晚年的演讲。而且，当有人将她作为演讲人来介绍时，她建议他们只介绍她是“第一台计算机的第三位程序员。”(有个不走运的人想做个更广泛的介绍，结果遭到了一顿海军式的斥责，这个人就是本文作者之一乔治·斯特劳)。我们同样关注 1994 年创办的格蕾丝·霍普女性计算大会 (Grace Hopper Celebration of Women in Computing conference)，大量女性 (现在男性也加入) 聚集在这一盛会上，解决 IT 专业人员中令人失望的性别不平衡问题。

### 对计算痴迷的数学家

霍普在瓦萨学院和耶鲁大学接受教育，于 1934 年获得数学博士学位，随后在瓦萨学院教了 10 年的数学。第二次世界大战期间 (1943 年)，她请假离开瓦萨学院，加入海军预备役，被分派至哈佛大学船舶计算工程局，与霍华德·艾肯 (Howard Aiken) 一起工作。1930 年代末，时任哈佛大学教授的艾肯说服 IBM 建造一台机电式计算机，取名为“马克一号 (Mark I)”。<sup>1</sup> (巧的是，马克一号仿照了 19 世纪巴比奇未完成的设计方案。) 1944 年，马克一号建造完成，交付哈佛大学使用。

如前文所提，霍普说自己是“第一台计算机的第三位程序员。”由于马克一号是机电式的，而不是电子式的，它没有被收入 IEEE 的 IT Professional 杂志 2014 年 3/4 月号的“电子数字计算机杰作”系列中。马克一号也许可以称得上是最后一台“前计算机”，可我们仿佛已经听见霍普对这番解释的责怪之声了。

战后，霍普没有回到瓦萨学院，而选择继续留在哈佛大学计算实验室。关于

她在哈佛的日子，人们经常会提到的一件事是她清除了马克一号线缆中的一只蛾子，这只蛾子破坏了马克一号的正常运行 (并导致其出现故障)。从那时开始，对软件和硬件的修正就被称为 debug (除虫)。<sup>2</sup> 1949 年，霍普加入埃克特计算机公司，成为通用自动计算机 (UNIVAC) 开发团队的一员。<sup>3</sup> 公司在 1950 年代早期陷入财政困难，被卖给了雷明顿兰德公司 (Remington Rand)。霍普继续在新的管理团队下供职，不久就作出了开创性的贡献，她开拓出了高级编程语言领域。

### 编译器与 Cobol 之母

霍普认为，计算机应该讲和人相似的语言，而不应该让人去讲计算机语言。她可能是持这一观点的第一人。1952 年，随着 A-0 编译器 (算术语言版本 0，它更像是一个链接加载器，而不是编译器) 的诞生，这个观点开始变为现实。<sup>4</sup> 1954 年，霍普被任命为 UNIVAC 部门自动化编程部门主任，她的部门发布了一些首次基于编译器的编程语言，包括 Math-Matic

( A-3 ) 和 Flow-Matic<sup>5</sup> ( B-0, 商用语言版本 0 )。其中, 1953 年发布的 A-2 编译器可能是世界第一款免费开源软件, 因为它允许用户免费使用源代码, 并鼓励用户提交改进建议。

不久, 人们意识到计算机编程是劳动密集型行业, 且耗资巨大, 开始准备开发和人的语言更相似的计算机语言。这种语言能够降低成本, 同时也会降低计算机程序的开发速度。这种语言的开发沿着几个方向进行。对于科学应用领域, IBM 于 1957 年开发了 Fortran( 即“公式翻译” ) 语言, 并推出了第一个用于 704 计算机的 Fortran 编译器。

对于计算机科学家来说, 由国际团队开发的 Algol 语言和第一个编译器诞生于 1958 年。面向人工智能研究人员的 Lisp 语言和用于 IBM 704 计算机的第一个 Lisp 解释器几乎与 Algol 语言同时诞生 ( Lisp 编译器直到 1960 年代早期才问世 )。不过, 不管是那时还是现在, 计算领域主要是面对商业应用。霍普将她的杰出才能都发挥在了商用方向。

霍普认为, 完成商用编程所使用的语言应该尽可能地接近英语, 这一观点尤其体现在 Flow-Matic 语言中。雷明顿兰德公司的高层对此观点表示怀疑, 但 1950 年代末 Flow-Matic 对用户发布, 立刻获得了成功。随后, IBM 用 Comtran( 即“商用翻译”, Fortran 的姊妹语言 ) 进入了商用编程语言领域。随着多种语言的陆续问世, 推出一部标准的商用语言的好处变得越来越明显。

对计算机应用程序重新编程的花销可能和首次编程一样高。比如, 购买一台新计算机就意味着要重新编写应用程序。所以在 1959 年, 一个由业内和政府人员组成的委员会得以成立, 其目标是开发

一套通用的、面向商业的编程语言。( 即 Cobol, 详见 <http://en.m.wikipedia.org/wiki/COBOL> ) 霍普在委员会中担任技术顾问。

Cobol 的第一个版本是 Cobol60, 多家计算机公司开始推出编译器。Cobol 获得了广泛的成功。1997 年, 研究机构 Gartner 集团估计, 现有的 Cobol 代码量总计高达 2000 亿行, 占全部商用程序的 80%。<sup>6</sup> Cobol 成功的原因之一是它的高度标准化。美国国家标准协会于 1968 年发布了第一部 Cobol 标准, 霍普是这一标准的主要推动人。霍普于 1967 年至 1977 年担任海军信息系统计划办公室海军程序设计语言团队的负责人, 并于 1973 年被授予海军上校军衔。

### “神奇的格蕾丝”

霍普曾于 1966、1971 和 1986 年三次从海军退役! 在她前两次退役之后均被军方召回, 并于 1983 年晋升海军少将。她获得了许多荣誉, 美国海军于 1996 年用她的名字为一艘驱逐舰命名 ( USS 霍普号 ), 美国能源部实验室将一台超级计算机命名为“霍普”。她的诸多贡献和荣誉让她获得了“神奇的格蕾丝”的称号。<sup>7</sup>

不久后, 她的退役演说又让她更富名望。她不仅是一位天才的演说家, 而且非常谦逊。她喜欢对听众讲关于自己的趣事: 有一次, 电梯里的一个人以为她是电梯操作员, 因为她穿着整套海军制服。霍普严肃纠正了这个人的印象, 而对方的回应相当无礼: “你一定是这儿最老的将军了。”霍普回击道: “不是! 里科弗将军比我还老!”

她的演说中另一个值得回忆的部分是她向在场听众发放的“纳秒”——一段

段不到一英尺长的电线, 这个长度大概就是光在 1 纳秒走过的距离。我们听她讲起在芯片诞生之前, 许多分立式晶体管的长度远超一英尺, 彼此间用电线相连。我们认为她的话是要让我们不要对 gigaop 计算机的诞生抱有希望 ( 芯片技术最终让 gigaop 计算机成为现实 ) 。



入计算机专业学习的女性人  
数很少, 而且越来越少, 这  
导致女性 IT 专业人员越来  
越少。高等教育群体正投入更多时间和  
精力来思考出现这一趋势的原因和解决  
方法。IT 行业需要国家 ( 和世界 ) 所能  
提供的每一位优秀人选, 所以将总人口  
的一半排除在行业之外是很糟糕的选择。  
榜样和英雄, 是鼓促人们选择某条道路的  
两大诱因。女性计算大会有助于这些目标  
的实现。用格蕾丝·霍普为这一盛会命名  
就是在向世界宣布, IT 英雄并非只有男性,  
女性同样可以成为 IT 英雄。■

( 本文内容来自 *IT Professional* )

## 参考文献

1. M.R. Swaine, “Harvard Mark I,” *Encyclopedia Britannica*; [www.britannica.com/EBchecked/topic/44895/HarvardMark-I](http://www.britannica.com/EBchecked/topic/44895/HarvardMark-I).
2. W. Isaacson, *The Innovators*, Simon & Schuster, 2014.
3. L.R. Johnson, “Coming to Grips with Univac,” *IEEE Annals of the History of Computing*, vol. 28, no. 2, 2006, pp. 32–42; doi: 10.1109/MAHC.2006.27.
4. P. Ceruzzi, *A History of Modern Computing*, The MIT Press, 1998.
5. Introducing a New Language for Automatic Programming: Univac

## 大师小传 MASTERMIND

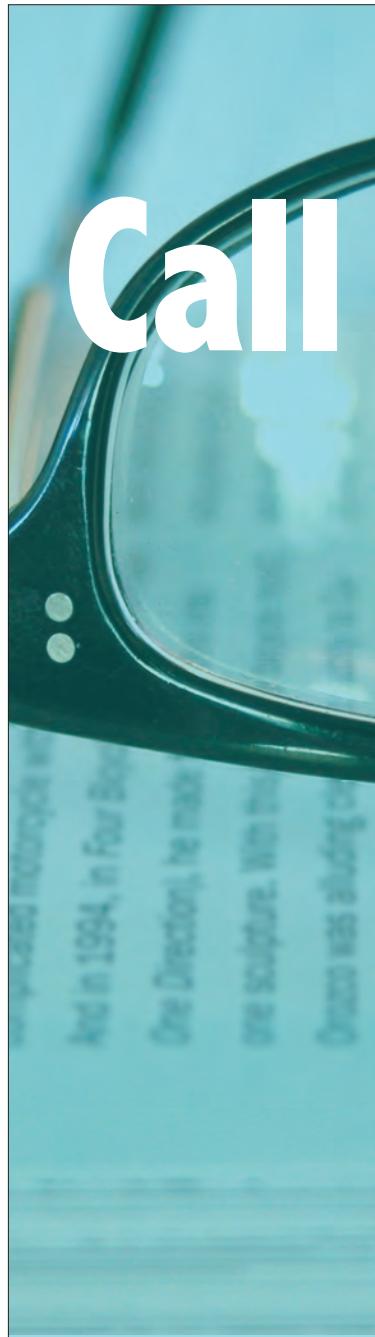
Flow-Matic, Sperry Rand Corp.,  
1957.

6. R.J. Kizior, D. Carr, and P. Halpern, "Does COBOL Have a Future?" Proc. Information Systems Education Conf., vol. 17, no. 126, 2000.
7. K.L. Engel, "Admiral 'Amazing

Grace' Hopper, Pioneering Computer Programmer," Amazing Women in History; [www.amazingwomeninhistory.com/amazing-grace-hopper-computerprogrammer](http://www.amazingwomeninhistory.com/amazing-grace-hopper-computerprogrammer).

乔治·施特劳 (George Strawn) 是网络与信息技术研发项目 (NITRD) 国家合作办公室主任。

坎迪斯·施特劳 (Candace Strawn) 是一位退休教师，曾在高中、社区大学和大学任教。



# Call for Articles 投稿

《IEEE软件杂志》希望刊登实用且值得一读的文章，可以吸引专业和非专业人士的兴趣。这本杂志的目标是把可靠、有用和前沿的信息传递给软件开法师、工程师和管理人员，帮助他们站立在技术浪潮之巅。文章主题包括需求、设计、构造、工具、项目管理、过程改进、维护、测试、教育和培训、质量、标准等等。

#### 投稿指南：

[www.computer.org/software/author.htm](http://www.computer.org/software/author.htm)

更多细节：[software@computer.org](mailto:software@computer.org)

**IEEE  
Software**



# CSP

CCF Certified  
Software Professional

# CCF软件能力认证

- 进入IT职场的通行证
- 参加认证达到一定水准，可获得企业优先聘用或高校计算机专业考研机考免试的待遇

2014年认证时间：9月21日、11月30日

## CCF会员享受特别优惠

Tel: 010-62562503-17/25  
Email: cspro@ccf.org.cn



登陆<http://cspro.org>报名参加



立即申请加入CCF

### 合作企业



HUAWEI



Baidu 百度



阿里巴巴 [Alibaba.com](http://Alibaba.com)



金蝶，企业管理专家



www.360.cn



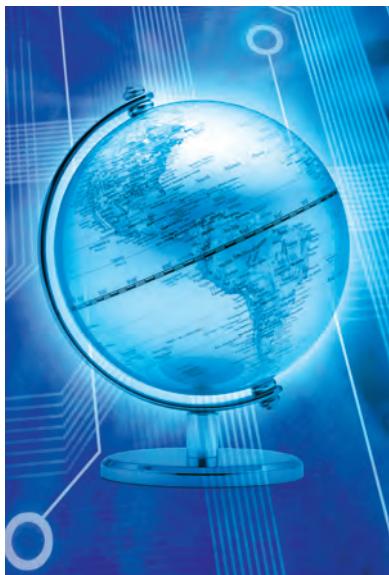
Tencent 腾讯



金山 KINGSOFT

### 合作高校





# 哥伦比亚的 人机交互教育： 建立产业和教育的桥梁

塞萨尔·A·科利亚索斯 (César A. Collazos)，哥伦比亚卡考大学 (Universidad del Cauca, Colombia)

路易斯·梅尔尚 (Luis Merchan)，圣布埃纳文图拉大学 (Universidad de San Buenaventura Cali, Colombia)

**H**人机交互 (human-computer interaction, 简称 HCI) 是一门与互动系统设计相关的、有影响力的学科。它出现在许多学科领域的交叉点，包括心理学，计算机视觉，人工智能，面部识别和运动跟踪等。近年来，改善人与计算机在所有方面的交互作用引起了越来越多的兴趣。然而，为了实现有效的智能化人机智能交互，计算机必须能自然地与用户进行交流，就像人与人之间的交流一样。<sup>1</sup>

先前的很多工作已经发现了人机交互教育中理论和现实的差距。<sup>2</sup>自从 1992 年《ACM SIGCHI 人机交互课程》(ACM SIGCHI Curricula for Human-Computer Interaction) 发表后，<sup>3</sup>计算机科学教育者用多种方法在选修课或计算机图形课、软件工程课、多媒体课，甚至是入门的计算机科学课 (作为某个关注的模块) 上应用了这些指导原则。然而，由于很多软件行业从业者不怎么了解人机交互，这些在本科和研究生教育阶段的人机交互课程在软件行业基本没有得到应用。

我们发现这点在哥伦比亚尤其典型。我们在哥伦比亚进行了一项调查。调查关于人机交互领域中学校教育和软件公司需求的差距。这项调查在哥伦比亚的一些大学和软件公司中进行，获得了很多重要的反馈结果，这些结果显示了在软件行业发展过程中人机交互的重要作用。

## 发现的问题

为了找出教育界和软件行业界之间的距离，我们就人机交互领域教育存在的问题对教授和软件开发者们进行了调查。我们尤其关注了如下几个方面：政府管理，软件业需求，以及课程设计和方法论。

## 政府管理

现在，强大的新兴科技已经可用于支持社会的可持续发展，使世界范围内所 有人受益。根据 2012 年联合国的一项调查，如哥伦比亚之类的一些发展中国家，在涉及政治的公民覆盖面和网络参与度上面已经开始赶上高收入国家。<sup>4</sup>因此对这些国家而言，加强获取用户信息和服

务诉求就成了势在必行的问题，尤其是涉及到在政府中民意代表很少的公民人群的需要和利益时。这些国家看待本国公民时，不能仅仅只将公众作为基于网络服务消息的被动接收者，还要鼓励他们成为积极的参与者，即基于 ICT 渠道传播与政府相关的信息而与政府进行互动——在这方面，人机交互的应用就变得非常必要了。

哥伦比亚政府的门户网站包含了很多帮助公众参与政府事务的功能 (见图 1)。公众可以使用多种工具，包括在线论坛、博客以及在线民调。这个门户网站还给用户提供了诸如 Facebook, Twitter, Wordpress, YouTube 及 Flickr 在内的社交网络参与渠道，让人们能够发表评论和观点。

在研究哥伦比亚官方政策的时候，我们发现该国信息技术和通信部给出的线上政府指导原则主要集中在官方网站的使用上。<sup>5</sup>这些政府提出的指导原则根据网站建设过程被分组，组别包括信息构建、用户交互设计、搜索，以及可用性和



图 1. 哥伦比亚的政府门户，上面有很多功能可以帮助公众参与政府事务

内容测试。每个指导原则都有一个相关的评估矩阵用来评价网站在是否符合国家、地方或地区的规定。

这个指导原则也列出了在网站中应用人机交互技术的好处：

- 减少生产成本（通过最小化后期设计变更次数），
- 减少支持和维护成本，
- 减少使用成本（通过降低使用难度实现），
- 更少的客户支持
- 减少培训成本。

然而，如何应用人机交互技术实现上述优点仍存挑战。

#### 软件行业需求

通过网上政府计划和信息技术部推

动的 IT 行业振兴计划，哥伦比亚政府正在努力提升 IT 产业在社会中的应用。政府项目为硬件和软件服务业提供了广泛的发展机会。从 2007 年到 2012 年，由于政府的行业振兴计划，哥伦比亚的软件业增长了 3.79 倍。哥伦比亚软件开发者协会 (Fedesoft) 与商务部、工业和旅游部一起，正在将哥伦比亚软件打造成出口商品。<sup>6</sup>

为了确定人机交互在软件业的需求潜力，我们调查了一些哥伦比亚软件公司。尽管软件业有很多人机交互的热点领域，但很少有软件公司在研究这些领域。在那些参与调查的公司尽管都对人机交互的重要性都有共识，但他们大多数多表示没有将人机交互手段集成到软件开发之中。所以我们需要一些方法以更好地实现学者和软件开发者之间的交流和信息共享，从而加强合作。

人机交互最有希望的研究领域之一就是在非传统领域的应用，比如移动端

或交互式数字电视。另一个重要的领域就是社交网络中的交互场景，而这正是社会中越来越普遍的现象。<sup>7</sup> 图 2 列出了软件开发中的几个人机交互热点领域。

#### 课程设计和方法论

哥伦比亚的大部分人机交互课程都是作为本科课程的选修课。<sup>8</sup> 其中涉及到最多的话题包括用户为中心的设计，可用性，以及交互作用设计。如图 3 中所示。然而，我们的研究也显示大学也正开始加入一些二级学科的选题，如无障碍性，人因体学，情绪设计，以及涉及到国际化和文化多元性等多种方面的因素。这项研究还显示应用最多的教学方法，是让学生将这些概念用到某个具体的案例研究中。

表 1 列出了哥伦比亚最好的 10 所大学，<sup>9</sup> 其中仅有一所将人机交互课程纳入了本科教学计划里。尽管诸如 ACM 的“2013 年计算机科学教程”(Computer Science Curricula 2013)<sup>10</sup> 等大量计算机科学培训训练参考标准一直在强调人机交互的重要性，并将其列为学科知识骨干体系，还把它纳入到计算机课程设计之中，但很显然涉及到人机交互的课程教学仍很不足。所以尽管硕士和博士项目都在推动人机交互，但在更多像“软件质量属性”这样的普通研究领域，人机交互还没有成为一个不可或缺的重要因素。所以在本科教育层次，人机交互课程并没有足够的广度和深度。

#### 建立产业和教育界的桥梁

用户满意度被普遍认为是质量管理的一项重要指标，一些研究显示它对于组织成本，利润和销售增长有着积极的影响。<sup>11</sup> 另一方面，用户的不满意则源于软

件开发过程中用户参与度的缺失。对于人机交互的情况而言，专家表示，用户通过反馈参与软件应用开发的整个生命周期，这对最后的成功非常重要。不过，系统专家们仍然不清楚需要多大的用户参与度以及人机交互如何帮助提高这种参与。

例如，可用性对于一个成功的软件产品来说是一个至关重要的指标。然而对于软件工程师克莱格·拉尔曼（Craig Larman）来说，当涉及到用户界面的可用性工程设计时，其应有的重要性和它在实际教育中受到的重视程度往往并不相称。<sup>12</sup>因此，我们需要考虑产业实际，将用户需求应用到人机交互的教学当中。一些整合方面的建议<sup>13</sup>对于特定的软件开发组织提供了特定的解决方案，但这些建议在其他不同特点的软件开发组织中缺乏通用性。

在软件行业开发产品时，我们需要教会人们意识到考虑人机交互机制的重要性。在巴西，一些公司通过咨询公司帮助他们进行产品可用性测试。在墨西哥、哥伦比亚和智利，一些公司也在产品可用性上努力，另外还有一些项目旨在整合学术界和产业界。<sup>7</sup>对于大学来说，将学术界和产业界的参与者联系到一起，进行一些考虑到产业需求的接地气的真实实验研究项目是非常有必要的，这些项目将会给学生一个将人机交互概念应用到产业实际的机会。另一个需要考虑的方面，是如何使人机交互的理念为软件开发者所熟知。

例如，可用性测试活动可以结合其它分析活动。将人机交互分析过程整合进其他通用分析过程相对容易。特别地，原型开发和产品设计这两个过程在人机交互中被视为“设计”活动，在软件工程里却被视为“分析”过程。<sup>14</sup>如果使用

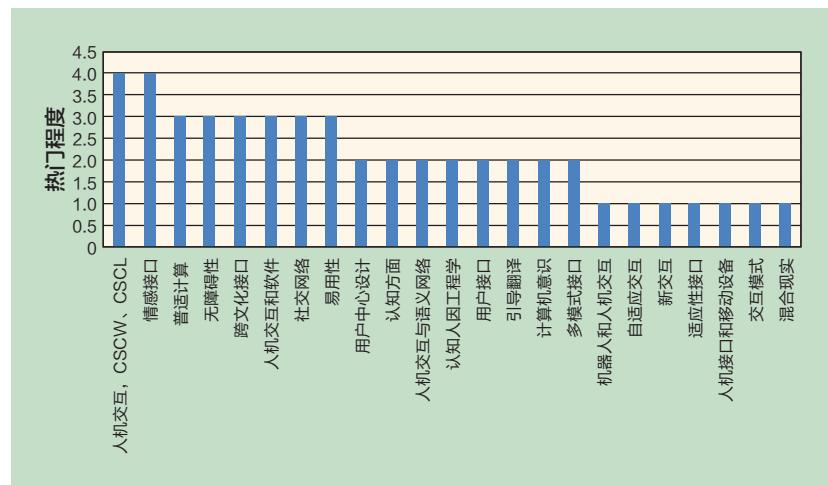


图2. 软件行业中人机交互领域研究的兴趣点。不幸的是，很多软件公司都不设计这些领域，这也是为什么人机交互研究人员和软件从业者需要更紧密的合作

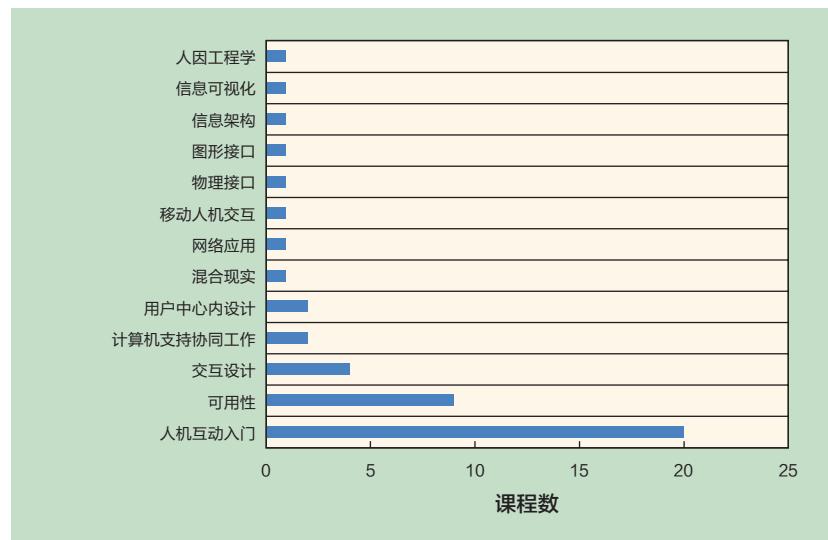


图3. 哥伦比亚的大学开设的人机交互课程

合适的术语和概念，在这两个过程中应用人机交互技术将会变得很容易。

我们还需要使软件开发者们意识到新的人机交互研究领域，让软件公司的开发者们受到更好的教育。在哥伦比亚，仅有0.1%的软件开发者拥有博士学位，1%拥有硕士学位，而这对于一个需要高水平劳动者保持竞争力的行业来说是相当低的。

人机交互领域非常多样，在以用户

为中心的一系列开发过程中没有一个通用的方法。不过大趋势似乎主要在于关注特定的使用环境，概述使用规格，开发原型，以及评估可用性。尽管对于像哥伦比亚这样的国家而言，大学主导的实验并没有像美国、英国和其它国家那样广泛，但随着研究者们越来越多地参与重要的研讨会，在有影响力的人机交互期刊上发表论文，以及在国际会议上受邀做主题报告，人机交互研究在哥伦比亚也开始

**表1. 无线心电图监测系统参数一览**

排名	大学	工程院系	可用性 / 人机交互课程
1	波哥大洛斯安第斯大学	计算机与系统工程	无课程
2	哥伦比亚国立大学，波哥大分校	系统工程	无课程
3	罗萨里奥大学	没有工程科系	无课程
4	埃克斯特那多大学	没有工程科系	无课程
5	伊赛斯大学	系统工程	无课程
6	麦德林行政、金融和技术大学	系统工程	无课程
7	哥伦比亚国立大学，麦德林分校	系统工程和信息学	无课程
8	沙巴纳大学	计算机工程	无课程
9	哈维里亚那大学	系统工程	有课程(选修课,人机交互,3学分)
10	北巴兰基亚大学	计算机与系统工程	无课程

慢慢进步和发展。随着哥伦比亚将人机交互技术活动带进主流软件工程活动中，软件公司将会提供更好的产品和服务。



## 参考文献

- N. Sebe, M.S. Lew, and T.S. Huang, "The State of the-Human- Computer Interaction," *Computer Vision in Human-Computer Interaction*, LNCS 3.058, 2004, pp. 1–6, 2004.
- E. Buie et al., "How to Bring HCI Research and Practice Closer Together," Proc. 28th Int'l Conf. Extended Abstracts on Human Factors in Computing Systems (CHI-EA), 2010, pp. 81–84.
- T.T. Hewett, ed., *ACM SIGCHI Curricula for Human-Computer Interaction*, ACM Press, 1992.
- "United Nations E-Government Survey 2012: E-Government for the People," UN, Mar. 2012; [www.un.org/en/](http://www.un.org/en/) development/desa/publications/connecting-governments-to-citizens.html.
- "Lineamientos y Metodologías en Usabilidad para Gobierno en Línea [Guidelines and Methodologies in Usability for Online Government]," Ministerio de Tecnologías de la Información y las Comunicaciones, 23 Aug. 2010; [http://paginas-web.univalle.edu.co/reglamentos/pasos/documentos/GEL108\\_CIN-TEL-Lineamientos\\_y\\_metodologias\\_en\\_usabilidad.pdf](http://paginas-web.univalle.edu.co/reglamentos/pasos/documentos/GEL108_CIN-TEL-Lineamientos_y_metodologias_en_usabilidad.pdf).
- Sector de TI en Colombia, tech. report, Federacion Colombiana de la Industria de Software, 2012.
- C. Collazos, T. Granollers, and M. Ortega, "Hacia una Integración de Interacción Humano-Computador en las Estructuras Curriculares a Nivel Iberoamericano [Towards Integration of Human-Computer Interaction in Curriculum Frameworks at the Ibero-American Level]," *Revista Internacional de Educacion en Ingenieria*, vol. 3, 2010, pp. 1–10.
- T. Granollers, C. Collazos, and M. González, "The State of HCI in Ibero-American Countries," *J. Universal Computer Science*, vol. 14, no. 16, 2008, pp. 2599–2613.
- "Mejores Universidades de Colombia, Según las Pruebas Saber Pro 2012 [Best Universities in Colombia, According to Saber Pro 2012 Tests]," Centro Virtual de Noticias de la Educación, 11 Sept. 2013; [www.mineducacion.gov.co/cvn/1665/w3-article-328609.html](http://www.mineducacion.gov.co/cvn/1665/w3-article-328609.html).
- "Computer Science Curricula 2013," Association for Computing Machinery, 20 Dec. 2013; [www.acm.org/education/CS2013-final-report.pdf](http://www.acm.org/education/CS2013-final-report.pdf).
- H. Jun, S. Kim, and C. Chung, "Measuring Software Product Quality: A Survey of ISO/IEC 9126," *IEEE Soft-*

# 新兴市场 NEW MARKET

ware, vol. 21, no. 5, 2004, pp. 88–99.

12. C. Larman, UML and Patterns: An Introduction to Object-Oriented Analysis and Design and the Unified Process, 2nd ed., Prentice Hall, 2001.

13. K. Radle and S. Young, “Partnering Usability with Development: How Three Organizations Succeeded,” IEEE Software, vol. 18, no. 1, 2001, pp. 38–45.

14. X. Ferre, N. Juristo, and A. Moreno, “Improving Software Engineering Practice with HCI Aspects,” Proc. 11th Int'l Conf. Software Engineering Research, Management and Applications (SERA), 2003, LNCS 3026, 2004, pp. 349–363.

**塞萨尔·A·科利亚索斯 (César A. Collazos)** 是哥伦比亚卡考大学

(Universidad del Cauca, Colombia) 计算机科学系的教授。联系方式: ccollazo@unicauca.edu.co。

**路易斯·梅尔尚 (Luis Merchan)** 是圣布埃纳文图拉大学 (Universidad de San Buenaventura Cali, Colombia) 计算机科学系的教授。联系方式: lmerchan@usbcali.edu.co。



《IEEE软件杂志》为软件从业者提供了前沿观点，专家分析和深刻的洞察，让他们跟上日新月异的技术变迁。这本杂志还为软件理论转化为应用提供了权威观点。

[www.computer.org/  
software/subscribe](http://www.computer.org/software/subscribe)

**SUBSCRIBE TODAY**



CCCF

从2005年3月创刊到2014年8月,CCCF已出版发行**102**期

期间我们刊登了

中国计算机事业创建50周年纪念大会专辑

中国计算机大会等多个专刊

云计算、物联网、大数据等几乎所有热点方向的专题  
以及脍炙人口的专栏、人物专访、动态、译文……



CCF会员可免费获得本刊  
欢迎加入中国计算机学会

# 中國計算機學會通訊

《中国计算机学会通讯》(Communications of the CCF, CCCF)是CCF旗舰刊物，2005年创刊，月刊。2014年3月正式推出iPad版。刊物设有：特别报道、专题、专栏、视点、动态、译文等栏目。邀请资深专家撰稿，也欢迎读者投稿。

# 医疗领域的 技术进步： 迈向个人化

阿尔夫·韦弗 (Alf Weaver), 弗吉尼亚大学 (University of Virginia)  
雷内·布赖斯 (René e Bryce), 北德克萨斯大学 (University of North Texas)

目前已经取得的技术进步和正在开发中的新技术随时准备在医疗领域掀起一场革命，这将为实时、个人化的病人监测和治疗带来大量机会，同时也给医疗数据安全带来了重大风险。

# 在

未来 5 到 10 年内，我们在医疗领域将会取得什么成就？随着我们对身体运行机制的理解越来越充分，由此产生的技术的迅速进步让医生和我们都能与我们的身体进行交流，交流方式是我们以前不可想象的。我们已经处在真正个人化医疗的边缘，个人化医疗对病人的具体特点进行个性化的关注，让每个病人都独一无二，以取代以前的“一刀切”的方法。

乳腺癌的确诊不再是问诊的结束，而仅仅是问诊的开始。对病人进行基因组测序能够让医生确定病人所患乳腺癌的种类，然后根据所得数据定制出有效的治疗方案，以获得希望的结果。同时，

是定制化的 Fitbit 及其竞争对手的产品 --- 也在进行大量的数据报告、收集和储存。

确实，移动设备的成熟程度已经达到了智能数据收集机器的程度，能够收集并分析生理学数据，然后将结果报告给可穿戴设备和任何远程观测者或监视者 --- 不管是人、软件还是二者的结合。用适当的方案设计的智能系统可以在医院内使用，也可以连接至远程位置作为设备和人员补充，成为现有的医疗基础设施的一部分。同时，廉价的健康监视设备和主要面向个人的健身追踪设备出现了爆炸式的增长。所有这些趋势都对新的安全要求和隐私控制作出了无可争议的需求：鉴于互联网具有无限的存储空间，

电子方式储存并管理的自己的健康信息。从数据的生成方式来看，EMR 和“限制”在美国医疗机构的 PHR 一般都在 1996 年的《健康保险携带与责任法》( HIPAA ) 的管辖范围内，该法案负责管理与 EMR 相关的隐私等问题。但是，面向设备的 PHR 不在 HIPAA 的法律保护范围之内，所以对这类 PHR 来说，要想获得广泛接受，病人就必须要信任那些对 PHR 数据进行收集、储存和传播的公司和系统。文中指出，将控制权从供应商转移到消费者手中不是件简单的事。PHR 数据披露产生的内在反应和风险可能会很严重，不论这种披露是有意还是无意。聚合的健康信息对制药商、保险公司、营销商和雇主等第三方具有很高的价值，而这些第三方不会总是为病人的最佳利益着想。系统架构师们，记住这一点！

移动设备及其在医疗管理中越来越重要的作用，是玛丽娜·维利科娃 ( Marina Velikova ) 、彼得·J·F·卢卡斯 ( Peter J.F. Lucas ) 和马尔藤·范德海登 ( Maarten van der Heijden ) 合作的《实现疾病自我管理的智能移动系统》一文的主题。今天我们的思想可以时刻与互联网连接，不久以后，我们的身体也会将个人的生理状态不断向记录、分析和预测软件报告。完全利用智能手机内部的仪器，比如利用麦克风来测量肺功能或利用手机摄像头确定血氧浓度，会为疾病自我管理带来新的机会。即将到来的定制化、嵌入式选择方案将加快技术进步的速度，同时也会带来更大的风险。新的硬件和网页及移动设备上的新应用会带来高效的个人化医疗，

## 实现真正个人化医疗的宏伟前景， 要从可靠、可访问且安全的 病人数据入手。

廉价的健康监测设备和个人健康记录的出现，让病人能对自己的全面健康状况的监测和管理起到更加积极的作用。

实现真正个人化医疗的宏伟前景，要从来自诸多信息源的病人数据入手。这些数据随后必须加以储存，确保其可靠、可访问、可共享，同时还要保证数据的安全。接下来就需要能够确保以上属性的系统架构。尽管这些系统还将继续“住”在医院和诊所等医疗企业环境中，但个人可穿戴设备 --- 无论是智能手机类，还

## 本期内容

李景全的文章《个人健康记录系统中的隐私保护》关注隐私和安全需求，本文在电子病历 ( EMR ) 和个人健康记录 ( PHR ) 之间划出了一条重要而又时常被人忽视的界线，前者基本上由专业医疗机构生成并掌握，而后者是个人以

## 关于作者

也会不可避免地产生新的问题。

服务质量 (QoS) 就是其中一个潜在问题。请考虑以下情况：

- 一位住院病人躺在术后台上，心律处于被监测的状态。
- 一位接受特别护理的病人，有多项生理学指标处于监测和分析中。
- 病人在家，或者通过不断收发数据进行远程诊疗，但没有固定的就诊安排。
- 通过移动设备所有这些活动进行监视的医生和其他医务人员。
- 必须仔细观测系统的总体表现的医院监测站。

面对这些不同情况，移动设备能够以实时方式充分收集和显示数据，并保证服务质量吗？在《实现医用级别的实时移动医疗服务》一文中，作者对参数空间进行了研究，系统可以在参数空间中提供“一定的数据传输速度、可靠性、隐私和安全”，为医院内和远程应用提供“实时、保密、准确的服务”。

由各方生成的数量庞大的医疗数据应该保存在哪里？应该如何共享？这些数据足够安全吗？使用这些数据可以达成什么目标？《医疗数据整合与云端信息学》一文的作者阿什迪普巴哈 (Arshdeep Bahga) 和维杰·K·马蒂塞蒂 (Vijay K. Madisetti) 提出了基于他们的原型云健康信息系统技术架构 (CHISTAR) 的新框架以及中间件，以协调云端数据分析，对来自不同的利益相关人的不同格式的

**阿尔夫·韦弗 (ALF WEAVER)** 是弗吉尼亚大学应用研究所的创所负责人，计算机科学教授。他的研究领域包括计算机网络、远程医学、电子商务、医疗数据隐私与安全和众包等。他在伊利诺伊大学获博士学位。现任 IEEE 会员，美国计算机协会国家讲师。联系方式：acw@cms.mail.virginia.edu。

**雷内·布赖斯 (REN E BRYCE)** 北德克萨斯大学计算机科学与工程助理教授，她的研究领域包括软件测试，对网页和移动应用风的特异性组合测试。布赖斯在亚利桑那州立大学获博士学位。联系方式：reneebryce@gmail.com。

医疗数据进行收集、整理和安全交换。网页和移动应用生成器可以让用户实现多种功能，如流行病监测、不良药品事件预测及医学预后预测等。

正当移动设备随时准备对个人化医疗进行革命性改变之际，普遍而价格实惠的全基因组测序 (whole genome sequencing) 有朝一日也许会实现更远大的目标：展现该测序方法的全新的优势与挑战。在《全基因组测序：革命性疗法还是隐私的噩梦？》一文中，作者埃尔曼艾蒂 (Erman Ayday)、艾米利亚诺·德·克里斯托法罗 (Emiliano de Cristofaro)、让·皮埃尔·乌鲍 (Jean-Pierre Hubaux) 和雷内·苏迪克 (Gene Tsudik) 解释了全基因组测序如何引领以“预测型、预防型、

共享型和个人化”为特点的 4P 医疗。但是，全基因组测序和其他一样，也是一柄双刃剑：一方面，它可以精确地定位患处并预测疾病的出现，使得病人可以在发病早期进行救命治疗。另一方面，基因图谱上的精确的生物信息学特异性可能导致个人隐私被无可挽回地破坏。



人化医疗已经成为现实，随着个人化医疗诊断和预测能力的提高，其影响力和影响范围都将不可避免地继续扩大。作为技术人员，我们的职责就是满怀热情与创新精神，竭尽所能地推进医学硬件和软件的开发，同时加强数据隐私保护。■





## 搜索你的工作机会

IEEE Computer Society 招聘可以帮你轻松找到IT、软件开发、计算机工程、研发、编程、架构、云计算、咨询、数据库很多其他计算机相关领域的新工作。



**新功能：**找出那些建议或要求拥有IEEE CS CSDA或CSDP认证的工作！

点击[www.computer.org/jobs](http://www.computer.org/jobs)，  
从全世界的雇主那里搜索技术工作岗位和实习机会。

<http://www.computer.org/jobs>

IEEE  computer society | JOBS

IEEE计算机协会是AIP Career Network的合作伙伴。其他合作伙伴包括《今日物理》杂志（Physics Today），美国医学物理协会（American Association of Physicists in Medicine），美国物理教师协会（American Association of Physics Teachers），美国物理学会（American Physical Society），AVS科学和技术学会（AVS Science and Technology），物理学生协会（Society of Physics Students）和Sigma Pi Sigma。



李景全 (Jingquan Li)，得克萨斯农工大学 (Texas A&M University)

个人健康记录是自行管理的医疗服务的组成部分，但记录的访问权由病人控制，这引起了人们的严重担忧，需要人们更好地平衡个人化、隐私保护和安全控制等要素。



人健康记录（以下简称 PHR）的出现让病人有机会对自己的健康信息进行储存、管理和共享，对病人的个人保健和护理有着广泛的影响。慢性病或重症病人可以记录自身的病情、症状及治疗过程，并和医务人员（如医师、护理师、精神科医师、实验室工作人员等）保持不间断的联系。对于病情相对较轻或间歇性发病的人，甚至是目前没有健康问题的人来说，个人健康记录在自主管理健康状况方面也具有不可估量的价值，自主管理健康状况可以降低医疗成本。在治疗中，个人健康记录可以增强病人与医疗机构的联系，能够提供的信息可能比传统治疗方式完整得多，在急救中，为病人施救的可能不是其主治医生，让这些信息更显得尤为重要。

越来越多的医疗服务商、保险公司和雇主认识到了这些有利之处，开始提供 PHR，结果造成在个人健康记录服务器上交换的数据量和数据类型越来越多。正如本文侧栏上所示，“PHR 包括什么？”，一份个人健康记录可以包括病人的医疗状况和病史、用药、精神健康、基因组成、性行为、生活方式、信仰和习惯等信息。其中的一些数据必须开放共享，以使病人获得适当的治疗，

但除此之外的其他数据则必须保密，因为在未授权的情况下披露这些数据可能给病人造成伤害。<sup>1</sup> 个人健康记录中的数据有很高的商业价值，使其成为不法商家、身份窃贼和腐败组织的诱人目标。

因此，个人健康记录系统在为个人化医疗管理提供了新机遇的同时，也带来了一些隐私和保密性上的风险。可以理解，病人担心自己的数据会被二次使用，这将使个人健康记录系统的信任度大打折扣，减慢病人对系统的接受度。为了扭转这一趋势，研究人员必须要面对技术上和法律上的双重挑战，防止未授权的数据访问和数据利用，这些未授权行为会导致雇主和保险公司的歧视、医疗身份窃贼和商业炒作等。<sup>2</sup>

对个人健康记录的利用相对较新，所以在解决信任和隐私问题时，还未研究如何利用 PHR 系统的架构属性来提供由病人控制的数据权限和安全保护。每种 PHR 系统都有其特定的优势，供应商可以将其合并成为一个混合架构，通过独立内容管理、独立隐私和安全审核以及法规遵循等隐私原则给予病人对其个人健康记录的唯一控制权。架构设计必须考虑隐私和安全风险，其中一些风险只和 PHR 访问权限有关。

表 1：比较个人健康记录（PHR）的性质

属性	限制性 PHR	非限制性 PHR	
		基于网络	基于设备
互操作性	不可互操作	可互操作	可互操作
访问通道	专用入口或客户服务器	互联网入口	电脑设备驱动
数据源	电子病历（EMR）	电子病历和消费者添加的信息	电子病例和消费者添加的信息
完整性	不完整	完整或部分	完整或部分
整合性	高	取决于用户提供数据的准确性和一致性	取决于用户提供数据的准确性和一致性
主要风险	可能无法向其他 PHR 系统传输，消费者可能不能获取数据。	服务供应商和商业伙伴可能把 PHR 信息作为商用或二次使用	物理损失、失窃、损坏及安全风险
法律规制	受 1996 年的 HIPAA 法案管辖	不受 HIPAA 法案管辖	不受 HIPAA 法案管辖
隐私控制	主要医疗网站控制数据	消费者和服务商控制数据	只有消费者控制数据
安全管理	安全外部网入口	可使用强加密和身份验证	可使用强加密和访问控制
安装或适用实例	梅奥诊所和凯萨医疗集团	Dossia 财团的 dossia.org 和微软的 HealthVault	CapMed 的 HealthKey 和 MedicAlert 的 E-HealthKEY

尽管供应商已经逐步采取措施，提供网络 PHR 系统将数据控制权给予病人或消费者，但许多隐私问题依然存在，比如怎样实施关于数据二次利用的规则。尽管如此，PHR 系统仍具有强大优势，激励着病人和 PHR 系统设计师们不断努力，打造出可以满足双方需求的产品。

限制性 PHR 系统都在 PHR 数据种类、数量和管理方式上给予病人更大的控制权，但在安全和隐私方面也更为宽松。

如表 1 中所示，互操作性、可达性、完整性、整合性和数据来源定义了 PHR 系统的运行特征，而主要风险、法律规制、隐私和安全问题则定义了 PHR 系统的局限性和可能面临的推广障碍。

由于每种 PHR 系统都有其优势和局限性，混合型 PHR 系统可能是颇具前途的新方式。

更方便地和医生分享信息，并在定期就医的间歇期更好地打理自己及家人的健康。这类系统的一个强大优势在于可控的安全和隐私，因为医疗服务商必须遵守 1996 年颁布的《健康保险携带与责任法》（HIPAA），这项联邦法律对个人健康记录的访问权限作出了严格限制。

尽管限制性 PHR 系统从表面上看似乎很理想，但它以机构为中心，会导致其存在重大局限性。因为医疗服务商控制着信息访问权限，不同机构的 PHR 中的数据量可能会有很大不同，而且数据既不便携，也非终身保存。<sup>3</sup> PHR 格式与主机构电子病历的格式相同，在大多数情况下，这会阻碍病人将 PHR 信息与其他医疗服务商和临床医生共享。当病人与机构的关系终止时，PHR 的访问权限也随之关闭。

对大多数医疗服务商来说，限制性 PHR 也并非热门的选择。因为他们在病人上门就医的间歇期内鼓励病人治疗收不到任何补偿。这种方式在逐步给予病

## PHR 系统的类型

医疗供应商、医疗计划和其他相关实体可以决定对 PHR 系统是否作出限制。表 1 中给出了每种 PHR 系统的一些架构属性。限制性 PHR 是医疗机构电子病历（EMRs）管理系统的扩展，用电子病历的内容来填充 PHR 的过程基本上是自动的，病人只有有限的访问权限。

而非限制性 PHR 系统则分为两类。基于网络的系统，个人健康记录储存在网页或云上；而基于设备的系统，数据则存储在智能手机等移动设备上。这两种非限

### 限制性 PHR

限制性 PHR 系统是一套闭合解决方案，由主机构控制访问权限和安全，向病人提供应用程序来访问门诊的电子病历。使用限制性 PHR 系统的医疗机构一般是优质机构，如梅奥诊所（Mayo Clinic），或者是有医保计划的机构，如凯萨医疗集团（Kaiser Permanente）。限制性 PHR 系统的目的是让病人访问自己的病历，

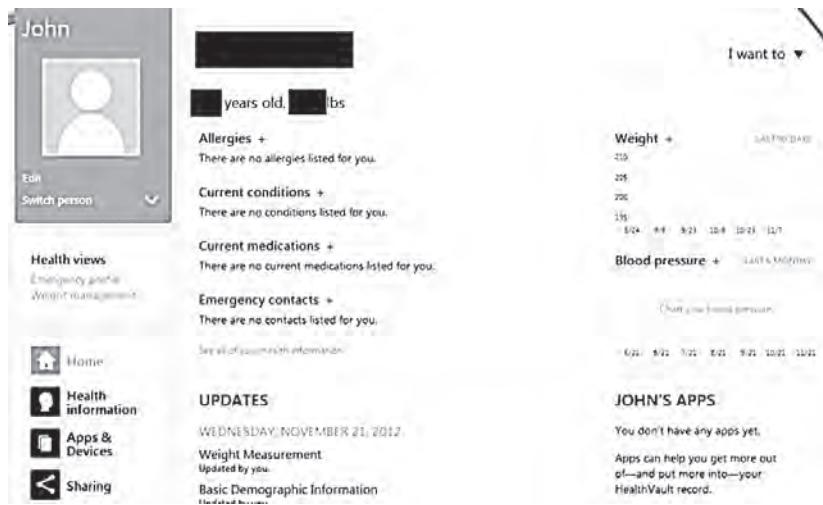


图 1. HealthVault 的主页。HealthVault 是一种基于网络的个人健康记录 (PHR)，在健康选择上给予消费者更大的控制权，用户可以自由选择记录中要包含哪些信息、和谁分享

人健康信息访问权，但移动医疗的趋势正在将这一需求朝着以消费者为中心的解决方案转变。

### 基于网络的非限制性 PHR

在 PHR 图谱的另一端是基于网络的 PHR 系统，由 Dossia Consortium ([www.dossia.org](http://www.dossia.org))、微软的 HealthVault、WebMD 和 PatientsLikeMe 等技术供应商提供支持。图 1 是微软的 HealthVault 的主页，上面给出了消费者在管理自己健康信息上的控制权限。

这种 PHR 很有吸引力，因为消费者对其个人健康信息有直接、可移植的访问权，可以将信息提供给其他一些医疗服务。和限制性 PHR 不同的是，基于网络的 PHR 让消费者来决定个人健康记录中包括哪些数据，可以与哪些人共享。医疗服务商可以通过全面获取病人的整个病历信息而获益，<sup>4</sup> 包括饮食和锻炼水平，这可能影响病人目前的状况。一些网络 PHR 系统，如 HealthVault 和 WebMD 可以免费下载。

可移植性是一个首要优势。比如，一位糖尿病人匆匆赶到一间急诊室就医，

即使急诊室不属于其最初就医的医疗机构，急诊室的医师仍然可以获得对治疗至关重要的信息。

与这些强大优势相对的则是隐私和安全上的隐患。HIPAA 法案中的隐私规定只对覆盖范围内的组织有效，包括医疗供应商、医保计划和医疗数据交换中心，<sup>5</sup> 并不包括技术供应商及其商业伙伴。如果信息从法案覆盖范围内的组织的文件上转移到了技术供应商的服务器上该怎么办？

尽管基于网络的 PHR 供应商保证让用户掌握数据的控制权，他们还是会介入到病人和最私密的数据中间。这些服务背后的商业模式将严重依赖于广告收入，以及与健康相关的第三方产品和服务的合作。这种依赖性会引发一些重要问题。

- 供应商追求更多利润的欲望，是否会导致其将用户的详细医疗数据共享给营销公司或众多合作伙伴？
- 消费者对自己的数据是否有独一无二的控制权？
- 如果消费者想要关闭账号，是否可以按自己的意愿处理数据？

对这些问题如果不能给出明确答案，会降低消费者的信任度。消费者可能相信公司提供的 PHR 处于隐私保护之下，但一旦他们开始收到第三方应用和服务，就会意识到许多非医疗企业也看到了他们的敏感信息。

### 基于设备的非限制性 PHR

与基于网络的 PHR 的极度开放相比，基于设备的 PHR 在个人电脑或移动设备上记录并储存数据。这些设备通常都装有密码保护、数据加密、数据导入规则和数据共享控制功能的软件。比如 CapMed 公司的个人健康钥匙 (Personal HealthKey) 和 MedicAlert 公司的电子健康钥匙 (E-HealthKEY) 等。一些基于设备的 PHR 产品还允许将信息复制到存储设备上，比如智能卡和 USB 闪存盘上。

比如，移动设备让消费者能够通过智能手机和平板电脑来访问基于网络的 PHR 系统。在急诊或预约就医时，医疗服务商可以通过任何无线网络来打开设备，立即获得病人 PHR 数据的访问权，并将信息显示在医疗机构的计算机上。

尽管有了软件保护机制，但隐私和安全保障仍然是个问题。实验显示，许多基于设备的 PHR 产品的加密和口令强度很弱。<sup>6</sup> 同时，存有 PHR 信息的设备也可能损坏或丢失，比如遭遇自然灾害或失窃，这会对消费者造成严重危害。而另一方面，当医疗供应商将病人的设备连接到存有其他病人敏感信息的计算机时，也可能造成危害。

## 个人健康记录包括哪些内容？

**2**009年2月1日，奥巴马总统签署了《美国恢复与再投资法案》(ARRA)，该法案为医生和医院使用健康信息技术提供高达数十亿美元的激励措施。法案中还明确指出，个人有权从执业医师处以电子格式获得自己的病历副本，并可以授权用自己选择的服务将病历保存起来。

这些病历是病人个人健康记录(PHR)的基础，PHR是一系列数据源组成的个人健康数据的电子资源。PHR系统让病人能够以电子化形式对自己的健康信息进行储存、管理和共享，而无须通过医疗机构保存的电子文件和复印件。比如，一位重症病人可能去过好几家医疗服务商看病，每家机构对该病人都有不同的疗法、健康史、实验室检查结果、用药和个人信息。值得注意的是，没有一家医疗机构能负责任地保证病人在就诊时的记录是完整的。

一份PHR包含图A中的全部或部分信息，包括药房用药记录、实验室检测结果、保险公司要求提供的数据和病人自己关于总体健康情况的记录，如血糖水平或体重变化等。PHR系统的首要目标是让病人有能力与授权方合作，在安全保密的环境下对健康信息进行访问和共享。

让病人自己掌握信息控制权是个人健康记录和电子病历的主要区别。只有受聘于医院和诊所等医疗机构中的专业医务人员才可以查看电子病历，而病人对自己的个人健康记录则有完全的访问权限，并可以决定将其共享给哪些医疗机构。



图A. 个人健康信息来源。个人健康记录可以包含多种数据来源，包括病人智能手机和平板电脑中的数据

### 参考文献

1. J.M. Grossman, T. Zayas-Caban, and N. Kemper, "Information Gap: Can Health Insurer Personal Health Records Meet Patients' and Physicians' Needs," *Health Affairs*, vol. 28, no.2, 2009, pp. 377-389.
2. "Connecting for Health: The Personal Health Working Group Final Report," 2003; [www.providersedge.com/ehdocs/ehr\\_articles/The\\_Personal\\_Health\\_Working\\_Group\\_Final\\_Report.pdf](http://www.providersedge.com/ehdocs/ehr_articles/The_Personal_Health_Working_Group_Final_Report.pdf).

实现对健康数据的最大化利用。

除在外。

### 混合型系统

为了克服限制性和非限制性PHR系统重大风险和局限性问题，一套灵活的PHR系统必须把前述两种系统的优点结合起来。比如，限制性或基于网络的PHR也可能允许消费者将PHR数据下载到U盘或个人设备上。混合型系统利用本地和远程存储产生PHR数据冗余，能够在任何环境下，以最灵活的连接方式

### 隐私和安全风险

在混合型PHR系统中对数据的一致性和完整性进行维护，需要在整个系统内保护PHR数据的机制。一款PHR产品的成功将严重依赖于其数据存取方案：个人健康信息必须让有合法需求的人方便使用，并将没有合法需求的人严格排

PHR系统面临的威胁包括由于病人对权限管理认识不清导致的泄露，从医疗服务商处的泄露，以及由于外部攻击和商用数据挖掘造成的泄露。

### 病人在访问权限控制上的失误

对消费者授权的二元性和日益增长的隐私风险会成为问题。医疗服务商长期控制着病历的存储和访问权，要想将控

制权交给病人绝不是简单的事。面对新的责任，病人会感到不知所措，可能会在授权或撤销访问权限时出现失误，导致信息的泄露或完整性的缺失。如果一些人缺乏通过网络或独立设备维护自己的健康数据的专业知识，就可能误删了对治疗至关重要的部分 PHR 数据。比如，PHR 系统设计不良和使用不当就可能导致病人在访问权限控制上的失误，危害病人安全或降低治疗质量。

### 医疗供应商的泄露

医疗供应商可能由于失误或故意泄露病人的信息。PHR 系统很复杂，医疗供应商可能会做出导致信息泄露的操作。内部人员可能出于恶意或为了获利而泄露病人信息。不管动机如何，这些信息的泄露可能造成严重后果：文献中明确记载了传染病病情、精神健康、慢性病诊断和基因组成等信息的泄露带来的风险。

PHR 中包含高度敏感的健康信息，这些信息可以透露出病人的身份。一旦医疗服务商泄露了 PHR 信息，病人基本无法将其恢复到私密状态。慢性病人对隐私风险的担忧尤其强烈，基本上不愿意他人访问自己的健康信息，而他们正是能够利用 PHR 获益最多的人。

### 外来攻击

通过互联网访问的 PHR 容易受到骇客和其他未授权方的攻击。同样，存有 PHR 信息的物理设备可能会被盗，导致银行信息和其他敏感数据的泄露。

一些公司在存储平台上提供基于网络的 PHR 系统，以商用和专利为目的挖掘个人健康信息。聚合化的个人健康数据对医药公司、营销商、保险公司和雇主都有很高的价值。

尽管在基于网络的 PHR 系统中，病人似乎可以完全控制信息的访问权限，但实际上他们对供应商及其大量合作伙伴如何使用这些信息几乎无法控制。对这些信息的一些二次使用可以获利，比如对处方药或健康监测中不良事件的监视。还有的利用这些信息从事营销诈骗，或者导致雇主和保险公司的歧视。

病人对这种不良性质的信息二次使用并没有什么反制的办法。2009 年颁布的《经济和临床医疗信息技术法》(HITECH) 大幅度提高了违反 HIPAA 法案带来的经济风险，并扩大了 HIPAA 中的一些规定，提高了对与商业机构联合违法的处罚力度。然而到目前为止，通过 HITECH 法案约束 PHR 供应商的只有联邦贸易委员会作出的一项规定，要求供应商提供某种对违法行为的提示过程，规定中并未对 PHR 供应商“合理使用或披露健康信息”做出定义。只有 PHR 供应商开始和法案管辖下的机构订立商业合作协议，这项关于健康信息使用和披露的规定才能适用。即便如此，如果 PHR 供应商宣布破产，之前的任何隐私协议都会作废。

### 目标：由病人控制的隐私策略

为了解决 PHR 系统独有的隐私挑战，PHR 系统供应商正在努力制定专门解决 PHR 信息披露和使用的正式策略。其中一种方法基于波士顿儿童医院的儿童医院信息学计划 (CHIP) 开发的个人控制的健康记录 (PCHRs) 平台，<sup>27</sup> 该平台能够让病人对自己的医疗数据的安全副本进行收集、维护和管理。

Indivo<sup>7</sup> 是现在世界范围内使用的网络 PCHR 实现方案，已经被 dossia.org 和 HealthVault 等网络 PHR 系统作为借鉴模型。Indivo 使用安全措施，将未授权方获得敏感的 PHR 数据访问权限的风险降至最低。

作为一种隐私保护策略，PCHR 解决方式的价值有限。因为病人和 PHR 供应商都可以控制 PCHR，其隐私情况取决于供应商的商业模式和商业利益。获得收入对于任何 PHR 产品的成功都至关重要，<sup>8</sup> 因此让市场的力量来决定 PHR 系统的性质和方向是不明智的。给予病人独一无二的访问控制权是阻止大多数 PHR 数据滥用的唯一方法。

其他隐私保护方式，如 W3C 隐私偏好 (P3P) 平台和企业隐私实践平台 (E-P3P)<sup>9</sup> 则试图阻止未授权的访问和数据利用。但如果病人没有更高的控制权，一些授权的个人或组织，比如 PHR 供应商及其商业伙伴同样可以为所欲为使用 PHR 数据。

这些访问权限和共享上的漏洞回避了问题的实质，“PHR 隐私是什么？”很简单，是病人对数据访问权的控制能力和确保 PHR 信息安全和完整的能力。<sup>10</sup>

在设计 PHR 系统时，让病人对自己的数据有完全的控制权，对于缓解他们对隐私问题的担心是很有帮助的。因此，由病人控制的隐私保护包括建立独立的同意授权管理，遵守监管规范，并在 PHR 系统内加入独立的隐私和安全审核。

### 同意授权管理

病人想要控制其 PHR 的访问权限和数据使用，就像控制自己银行账号里的资金一样。确保个人对信息访问和使用的控制权，可以增强消费者和技术供应商之间的信任，但如果允许技术公司掌握如此敏感的信息，会使一些病人感到不快。在某种程度上，他们意识到自己对数据并没有完全的控制权。

对于许多病人来讲，拥有将自己的个人健康信息授权作任何用途的权利，是隐私的根本所在。建立独立的同意授权管理，把访问电子数据的同意授权和 PHR 供应商相分离，是确保病人拥有这一权利的一种方式。除非经病人明确同意，否则连 PHR 供应商也不能将病人的个人健康信息共享，或用作其他非医疗目的。电子化同意授权管理必须作为任何消费者控制的 PHR 系统基础架构的组成部分。在系统设计中嵌入同意授权管理，可以确保 PHR 供应商在获得任何与病人有关的数据之前必须要经病人同意。<sup>11</sup> 病人会更有信心，因为他们对自己的医疗数据享有完全的控制权。

### 自动化的监管规范

并非所有的病人都具备足够的知识

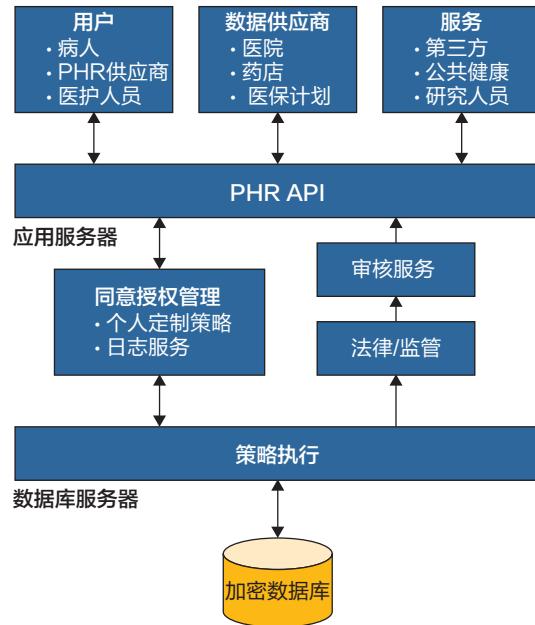


图 2. 由消费者控制的 PHR 系统架构。表现层 (API) 用来帮助病人识别隐私策略，应用服务器层使用隐私策略来控制数据的存取。策略执行层负责执行用户定义的隐私策略，将每条数据请求储存在日志文件中

和技术来维护自身健康数据的机密性。即使加入知情同意规则，一套符合监管规范的默认隐私和安全策略，以及公平的信息处理也会大大提升系统的可信度，进而增强信任，扩大系统的使用范围。

对每次 PHR 数据的使用和披露都进行审核，能够让病人确定哪些人或机构访问或修改了自己的 PHR 数据。PHR 系统必须保留对每次 PHR 数据披露和使用的审核记录，包括访问或使用数据的人和机构，以及这些行为的发生时间和目的等。独立的验证和审核服务可以确保 PHR 系统确实执行了病人创建的隐私策略。病人可以通过查阅 PHR 账户访问列表来监测违规行为，并采取相应行动。

对数据使用的监测和审核，是能够决定保护隐私的 PHR 系统质量和诚信的实用解决方案。比如，微软公司承诺的独立第三方审核，就为 PHR 系统的隐私保护设立了一个新的标准。独立的隐私和安全日志以及审核机制，对病人是否

信任和接受 PHR 系统起着决定性作用。

### 实现架构

大多数基于网络的 PHR 系统使用图 2 所示的三层架构，分别是表示层 (API)、应用服务器层和数据库服务器层。

**表示层。** 表示层负责处理数据访问请求，向数据用户和服务呈现信息，帮助病人指定应用服务器层和策略执行层使用的隐私策略。表示层试图作出的任何动作都会通过应用服务器层，因此这些动作会遵守病人选择的隐私策略。

**应用服务器层。** 应用服务器层是病人控制隐私策略的关键，负责处理医疗机构治疗病人需要的广泛信息中的商业逻辑，并将隐去病人身份后的数据传送给商业和研究机构。应用服务器层比基于网络的 PHR 系统多了三项服务：同意授权管理、策略执行和记录与审核。

消费者利用同意授权管理来控制 PHR 的访问权和数据的使用。即便有明确同意，隐私策略也必须符合监管要求，并易于审核。

顾名思义，策略执行服务负责实施隐私策略，隐私策略由病人制定，解决由所有利益相关人带来的风险，包括消费者、专业医务人员、PHR 供应商及其商业伙伴，甚至包括病人自己（病人可能会无意间违反隐私策略）。

每次访问请求都会生成一次事件，从执行服务传送至日志服务，后者负责将每一次数据披露和访问行为保存在日志文件中。审核服务会显示访问请求的日志文件，以证明 PHR 系统执行了隐私政策。

**数据库服务器层。**数据库服务器层负责保存 PHR 数据，数据经过加密，以便在数据泄露、骇客攻击和硬件失窃时保护病人。加密密钥保存在单独的物理服务器上，以防数据库服务器一旦遭到破坏，健康数据会被解密。服务器存储和数据网络加密共同为另一个保护层服务。

几乎没有人会怀疑 PHR 在引领我们走进移动医疗和病人控制病情管理权的时代。然而，隐私问题仍然是阻碍 PHR 系统广泛应用的威胁。PHR 供应商们想要靠承诺给予病人控制权来获得声誉。不论供应商的动机如何，隐私风险仍然是 PHR 系统全面推行的障碍。

尽管让病人获得唯一的 PHR 的访问权和数据使用权是极其困难的事，但这

却是确保 PHR 数据获得正当使用的最佳方式。由病人控制的隐私保护架构方案，以基本隐私原则为基础，催生出具备独有特色的 PHR 系统，而这只是解决方案的一部分。其他部分还包括详细的技术设计，管理实施机制，以及对消费者进行有关 PHR 的好处和数据隐私管理的教育。

另一个关于 PHR 研究的关键问题是，如何调整 PHR 供应商的商业利益以及对保护隐私及 PHR 数据机密性的奖励措施。PHR 供应商在保障病人隐私需求上的投资，需要获得收入或其他回报作为平衡。这种让供应商和病人双方都获得优势的平衡，是 PHR 系统广泛应用的关键。C

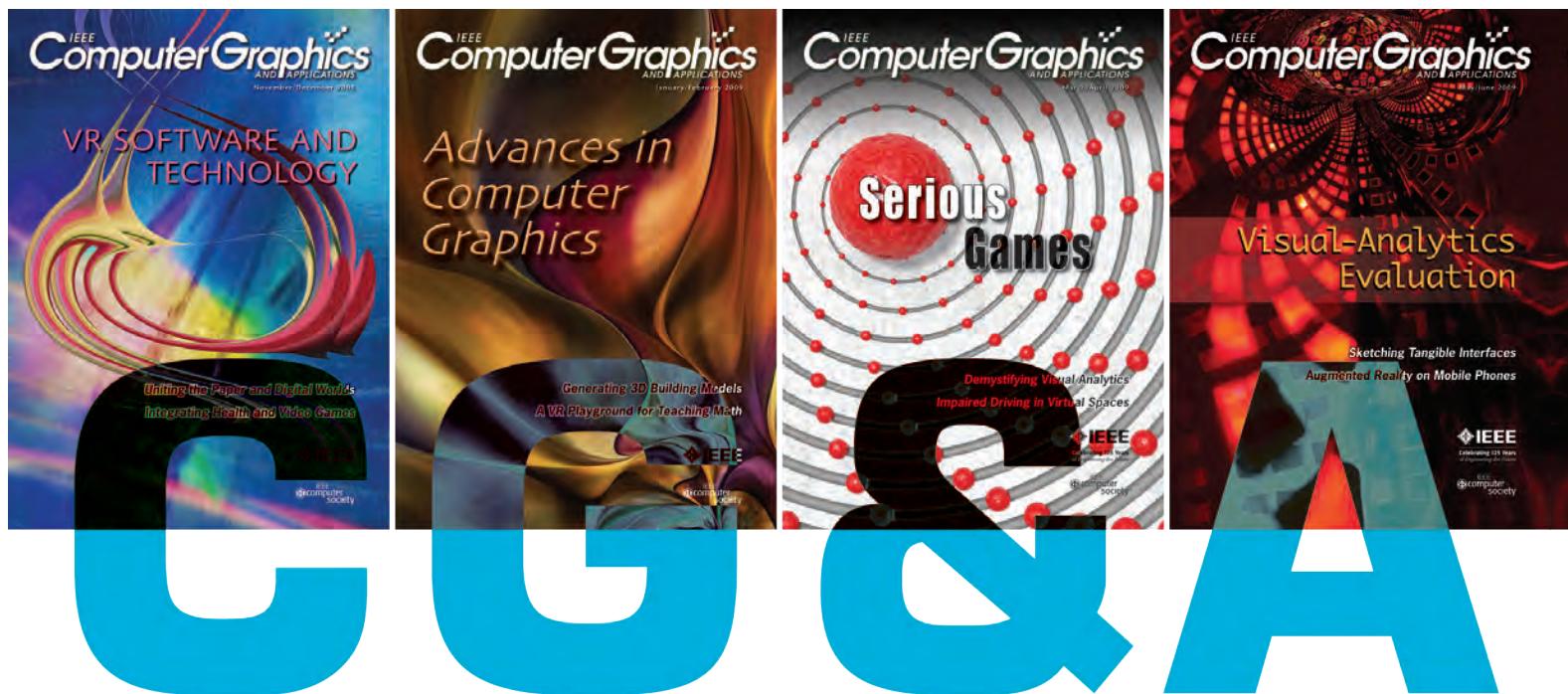
## 参考文献

1. T. Rind eisch, "Privacy, Information Technology, and Health Care," Comm. ACM, vol. 40, no. 8, →1997, pp. 93–100.
2. J. Li, "Privacy Policies for Health Social Networking Sites," J. Am. Medical Informatics Assoc., vol. 20, no. 4, 2013, pp. 704–707.
3. J.M. Grossman, T. Zayas-Caban, and N. Kemper, "Information Gap: Can Health Insurer Personal Health Records Meet Patients' and Physicians' Needs," Health Affairs, vol. 28, no. 2, 2009, pp. 377–389.
4. N. Archer et al., "Personal Health Records: A Scoping Review," J. Am. Medical Informatics Assoc., vol. -18, no. 4, 2011→, pp. 515–522.
5. J. Li and M. Shaw, "Protection of Health Information in Data Mining," Int'l J. Healthcare Technology and Management, vol. 6, no. 2, 2004, pp. 210–222.
6. A. Wright and D.F. Sittig, "Encryption Characteristics of Two USB-Based Personal Health Record Devices," J. Am. Medical Informatics Assoc., vol. →14, no. 4, 2007, pp. 397–399.
7. K.D. Mandl et al., "Indivo: A Personally Controlled Health Record for Health Information Exchange and Communication," BMC Medical Informatics and Decision Making, vol. 7, no. 25, 2007, pp. →1–10.

## 关于作者

李景全 是德州农工大学圣安东尼奥分校的助理教授。他的研究领域包括信息安全与隐私和数据挖掘等。他从伊利诺伊大学香槟分校获信息系统博士学位，现任 IEEE 计算机学会会员。联系方式：jli@tamus.tamus.edu。

8. L. Martino and S. Ahuja, "Privacy Policies of Personal Health Records: An Evaluation of their Effectiveness in Protecting Patient Information," Proc. 1st ACM Int'l Health Informatics Symp. (IHI '10), 2010, pp. 191–200.
9. G. Karjoth, M. Schunter, and M. Waidner, "Platform for Enterprise Privacy Practices: Privacy-Enabled Management of Customer Data," Privacy Enhancing Technologies, LNCS 2482, 2003, pp. 69–84.
10. "Guidance from CDC and the US Department of Health and Human Services," Centers for Disease Control and Prevention (CDC) HIPAA Privacy Rule and Public Health, MMWR Morbidity and Mortality Weekly Report, May 2, 2003, vol. 52, no. S-1,
11. S. Haasa et al., "Aspects of Privacy for Electronic Health Records," Int'l J. Medical Informatics, vol. 80, no. 2, 2001, pp. E26–e31.



《IEEE计算机图形及应用》(IEEE Computer Graphics and Applications, 简称CG&A)把计算机图形学领域的理论和实践联系在一起。《IEEE计算机图形及应用》提供了包括从某个特定算法到全系统实现在内的同行评议的深度报道。它为那些处于计算机图形技术前沿的人们提供了必不可少的资料。无论他们处于商界还是艺术界,这本杂志都能让他们受益。

请点击: [www.computer.org/cga](http://www.computer.org/cga)



玛丽娜·维利科娃 (Marina Velikova)，荷兰应用科学组织 嵌入式系统创新研究院  
彼得·J·F·卢卡斯 (Peter J.F. Lucas)、马尔藤·范德海登 (Maarten van der Heijden)，荷兰奈梅亨大学

廉价的移动医疗不仅要注重技术性能，  
而且要考虑病人和医疗供应商之间的责任划分，病人的病情背景，  
以及如何实现对病人决策的支持和定制化交流互动。

**使**用移动技术把各类医护人员（包括医生、护士、护理师和理疗师等）的医疗服务支持传给病人，从而实现对疾病的自我管理，是很有前途的方式。在向病人提供信息、参与治疗决策、监测病情以及针对意外变化对医护人员提供警告上，都有着无限的潜力。<sup>1</sup>

不过，要让病人持续活跃地参与移动医疗需要具备很强大的大局观，将疾病的自我管理视作一个过程，这需要对病人进行角色转换。如图1所示，在传统医疗中，病人在很大程度上是被动的，但在移动医疗系统中，病人会积极参与到关于治疗的决策过程中来。如此高的参与度需要病人掌握与疾病症状相关的足够知识和解决问题的技能，以产生行为变化和应对策略。活跃的每日自我管理包括定期监测并报告体征和症状、坚持用药，并按健康状况的变化作出适当反应。上述这些要求的内在核心是医患之间的密

切合作。

目前的大多数研究强调的重点都是移动医疗的技术，以及以病人为中心的其他形式的医疗，<sup>2,3</sup> 或者是专门的医疗应用。<sup>4</sup> 而我们则建立了两套移动医疗系统：一套用于慢性阻塞性肺病（COPD）<sup>5</sup> 的病人，另一套用于妊娠并发症的病人，目的主要是探索移动技术是如何为病人提供支持的。我们的重点是在智能手机中建立一个疾病专用的贝叶斯网络，对病人数据进行实时解读。这样，系统可以访问病人的健康状况，为病人提供行动建议，而这些过程都没有医疗提供商的参与。

实验显示，这两套移动医疗模型都产生了积极效果。用于慢性阻塞性肺病的系统是我们结合临床数据，与临床专家密切合作开发的，在实验中仅靠病人测量的生理信号和所报告的症状，正确检测出了91%的慢性阻塞性肺病的发作。用于妊娠并发症的

系统由我们与妇产科医生合作开发，在测试中，根据实际的妊娠数据，<sup>6</sup>在怀孕病人实际确诊四周前就可靠预测出了60%的高血压并发症。

作为研究成果的一部分，我们把为病人提供自我管理服务的移动医疗分为四个基本部分，对共同医疗的支持、背景感知、嵌入式智能和个人化交流，并探索如何将这些内容以最佳方式整合到移动技术之中。

病人自我管理面临的一大挑战就是确定最优复杂度，对特定背景下的临床数据进行展示和解读，以支持病人定制的交流、决策和反应。我们认为，我们使用的整合方式向着高成本效益的移动医疗系统迈出了一步，这种系统能够补充甚至替代传统的疾病管理，让更多的人以更低的成本享受更健康的生活。我们同样相信，这种整合方式将会促进面向病人的决策支持移动系统的科学、技术和商业的发展。

## 对共同医疗的支持

移动医疗疾病自我管理智能系统的成功，取决于病人和医护人员共同承担医疗责任的效率，嵌入式智能系统必须同时进行数据解读和反馈，并支持医患之间的交流和决策。

数据传输速度是一项关键共同医疗服务，因为实时数据传输对于监测病人的病情是必不可少的。在个人化移动医疗系统中，将由病人决定是否传输数据，并确定传输量和频繁程度。如果定期传



图1. 传统的疾病管理与病人疾病管理。在传统的疾病管理模式（绿色）中，病人被动地接受治疗方案，但在支持病人自我管理的移动医疗系统（黄色）中，病人主动参与病情监测和治疗决策，医护人员会收到关于病人情况的持续更新信息。系统可根据病人的偏好来定制交流方式，可以选择语音、短信、视频或电子邮件

输的数据意外停止更新，医护人员可能会分步检查传输的停止是否出于故意，并研究是否有必要介入。

另一项对共同医疗提供支持的服务是通过语音、视频、短信或电子邮件实现医患间的直接交流，以便及时讨论可能出现的问题和治疗方案的调整。病人可以主动联系医务人员，系统也会对是否有必要联系给出建议，从而降低了交流的频度和成本。这种定制化交流会促进医患双方积极参与疾病管理。

一套支持共同医疗的移动医疗系统很可能促进病人自己管理疾病，并可能产生比传统治疗方式更好的总体结果。比如，让病人对自己的临床信息拥有更便利的访问权，能够使他们更加积极地研究循证性临床指南和相关的科学研究，获得关于疾病治疗的更多知识。

## 背景感知

由于个人化是移动医疗系统的重要优势，移动设备必须能够对病人、病情和环境背景的变化做出正确的识别和解读，并适应情况的变化。

## 病人背景

病人的年龄、性别、个人和家庭病史，以及饮食、饮酒、吸烟和日常活动等生活方式，都会影响病人感染疾病或使已患病的病情变化的风险。移动设备可以方便地收集这些病人的特异性信息，比如可以让病人填写调查问卷来获取病人健康记录之外的信息。

个人因素与人际因素，如先前的健康相关行为、社会经济地位以及社会态度和支持也会影响改善健康行为和自我治疗的可能性。这些因素决定了病人的

生理和心理状态，而这两者又分别决定了一个人目前和未来的健康。健康状况以三种类型的数据为基础：

- 症状：由病人报告的主观体验，如咳嗽、乏力和头痛。
- 体征：对测试和测量设备得出的生物医学数据，如通过血压计测得的血压，由生化检测得出的血糖和血

有按流程操作，或是不具备必要的技能。比如，一位肾功能受损的老年病人可能很难快速分析尿液试纸上的颜色，而这就可能导致其错误报告这项检测结果。

妊娠高血压（gestational hypertension）和先兆子痫（pre-eclampsia）只需要在孕期的疾病管理。

## 为了支持疾病自我管理，必须向移动设备植入充足的信息，以作出智能化的健康决策。

通过移动设备上的调查问卷很容易收集症状报告，而体表传感器等测量设备可以收集体征和生物信号，并通过无线网络将其传输至计算设备。这些测量设备小巧廉价、携带方便、易于使用，测量结果比医院和诊室环境下获得的数据更具代表性。远程测量消除了“白领效应”，即病人在医院或诊室的状态会发生改变，同时降低了成本，因为病人去医院就医的次数减少了。

但另一方面，确保这种远程测量的可靠性可能会更加困难，因为病人可能没

有按流程操作，或是不具备必要的技能。

**长期疾病管理。**研究显示，有 75%-85% 的医疗费用花在了慢性病的管理上，如慢性阻塞性肺病、高血压和 I、II 型糖尿病。<sup>7</sup>许多慢性病已被充分研究，且可以预防，因为这些病与病人生活方式的选择密切相关。比如，90% 的慢性阻塞性肺病是由于病人长期吸烟而导致的。很明显，对该病的治疗要从戒烟开始，这使慢性阻塞性肺病成为评判移动医疗系统在处理与生活方式选择相关的长期疾病管理上效果好坏的杰出指标。

**短期疾病管理。**有些疾病不需要无限的管理，比如对许多运动损伤的治疗只需要几周时间。与怀孕相关的失调症，如

**重症管理。**重症管理主要是长期管理。以慢性阻塞性肺病为例，该病按轻微至严重分为几个阶段，而处于哪一阶段决定了对该病进行自我管理的合适程度和可能性。移动医疗解决方案会针对某一阶段向病人提供具体行为的适当建议。处于严重阶段的病人可能会经常收到用药指导和提醒，提示病人联系医护人员。而处于轻微阶段的病人只会偶尔收到提醒，提示病人测量一些指标，对健康状态进行评估。

## 环境背景

现有的移动设备内装有内置传感器来收集关于用户环境的数据。加速度计和三轴陀螺仪可以识别出运动和方向，麦克风会捕捉声学数据。最近关于这些内置传感器的研究表明，现在已经有些新应用能够利用智能手机麦克风来测试肺功能，<sup>8</sup> 或使用手机摄像头来测量血氧饱和度。<sup>9</sup>

移动设备与环境技术之间的通信，比如智能家电和可穿戴式传感器可以进一步帮助被隔离的病人保持健康和安全的标准。

## 嵌入式智能

尽管收集关于病人及其所处环境的数据的技术难度不断降低，但如何对这些数据进行解读，以支持疾病自我管理仍然

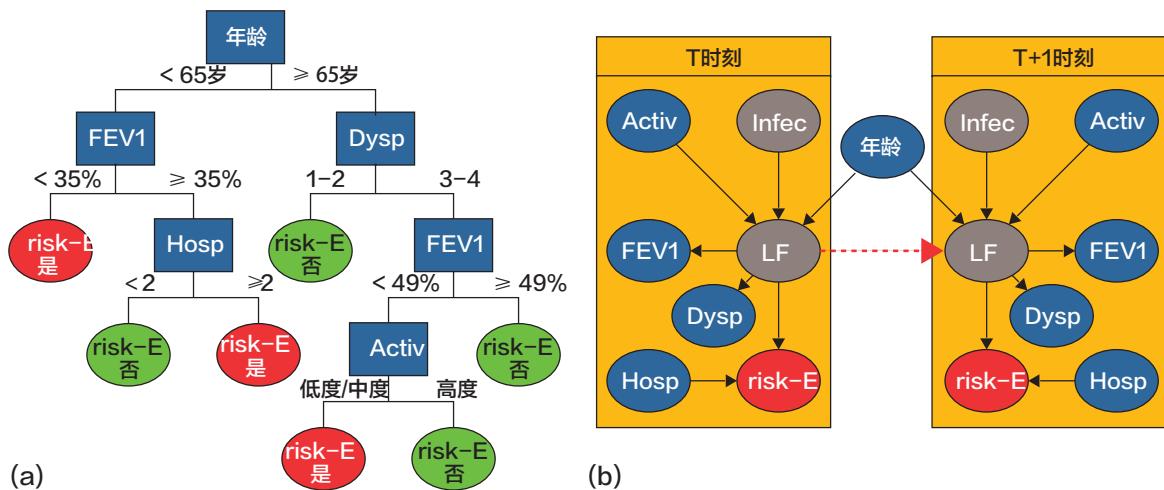


图2. 预测慢性阻塞性肺病恶化的两种模型。(a)决策树(b)时间贝叶斯网络。(b)中的实线表示时间片段中的因果相关性,虚线表示器官功能状态随时间变化的相关性。Activ: 活动; Dysp: 出现呼吸困难; risk-E: 有恶化风险; FEV1: 第1秒用力呼气量; Infec: 出现感染; Hosp: 过往住院史; LF: 肺功能

是个难题。对数据的解读必须考虑病人病情的临床背景,这即便对于经过广泛训练的临床医生来说都具有很强的挑战性。即便是具备人工智能算法的医疗决策支持系统,也只能协助管理一两种疾病。对范围广泛的一系列疾病仍然不能可靠地处理,因为疾病的相互作用具有很高的内在复杂性。在上世纪90年代研究人员发现如此大范围系统的准确性不足以,到现在为止在准确性上还没有什么大的进展。<sup>10</sup>

因此,要想对疾病管理系统提供足够的支持,必须在移动设备内植入可由病人操作的决策帮助,使病人获得足够的关于可用选择和预计结果的信息,做出明智的健康决定。这类决策帮助的精确的问题解决能力取决于以下几点:

- 对决策复杂性的需求程度
- 对医疗原因不明情况的处理办法
- 对特定的临床背景下的病人健康状况模型的选择以及
- 从病人指定的数据中学习个人化特征的方法

### 复杂性权衡

简单的决策帮助大部分基于假定规则,常以脚本语言编码。当所需数据(临床数据、位置数据等)缺失,或病人的某些实验室检测结果超出正常范围时,规则就会发出警告。

慢性病人不大可能符合人口学规范,慢性病人的个人化疾病管理需要一套单独的、高度复杂的规则集,很容易造成使用不便。可以利用决策树简洁地呈现这些规则,如图2a所示,这种管理方式可在移动设备上实现。

决策树用简单直观的形式实现了个人化决策,而且易于对非专业人员解释。但它不能确定疾病预后趋势,而预后趋势决定了病人未来的治疗方案。

### 对不确定性的处理

作为决策参考依据的医学知识和数据具有固有的不确定性,这意味着决策树等确定性方法在解决医疗原因的问题上不如概率性方法(如贝叶斯网络)和逻辑回归。<sup>11</sup>贝叶斯网络可以解决多个因果

关系,比如存在于症状、体征、疾病中的因果关系,并用概率分布对不确定性进行量化。贝叶斯网络能够解决复杂性和不确定性问题,<sup>12, 13</sup>成为在移动医疗系统中构建决策模型的很有前途的选择。

贝叶斯网模型描述了变量间的大致关系,建模者可以通过输入特定病人的数据来做出个人化预测,输入虚拟的证据来执行假设场景的分析,还可以利用贝叶斯网络来对病人特征的概率分布进行微调。

贝叶斯网络既适合没有数据可用的情况,也适合有大量数据可用的情况。对于前一种情况,建模者可以利用专业知识和文献信息;对于后一种情况,建模者需要依靠机器学习。最终,当可用数据只由几百例病历组成时,采用典型的拟合优度统计学方法(如逻辑回归)更适合贝叶斯网络构建预测模型。

### 静态与动态健康状态捕捉

支持病人自我管理的智能移动健康系统可以捕捉病人健康状态的静态或动态数据。在静态捕捉中,特定时刻组织功能决定了实验室检测的结果和某些症

状是否出现。凭静态捕捉基本上足以进行诊断。

不过，预后任务就需要能够在分析健康状态时考虑时间推进的动态模型，该模型可以做出预测，供疾病自我管理系统用来实现治疗方案的定制。动态模型必须考虑病人的特征，如已患疾病、年龄、受基因影响的器官功能以及使用治疗药物来治疗或减缓病情的发展。

## 预后工具需要能够在分析状态的同时 考虑时间推进因素的动态模型。

动态捕捉对于病人自我管理系统非常重要，因为它可以让病人和医护人员主动采取补救措施。如图 2b 所示，贝叶斯网络模拟出了在特定时间段内影响慢性阻塞性肺病恶化的一系列因素间的相互关系。该模型按照临床实践，将器官 X 的先前的检查结果（时间点 1 至 t-1）与 t 时间点的检测结果进行了合理比较，以确定器官功能是否发生了重大改变。系统能够根据历史和目前的状态，预测出器官的功能情况和随后的时间段内（时间点 t+1 至 T）病情的潜在发展。

### 个人化学习

移动医疗系统能够持续收集病人和环境数据，这使得人们可以提取个人化数据来定制模型结构，对统计模型参数进

行微调或更新，以便将病人的行为和病情的变化考虑进去。近期的实例包括基于决策树或支持向量机的模型，后者能够识别病人的活动，警告病人潜在危险，或建议病人选择更健康的生活方式。

建模者可以收集众多病人的数据，并利用离散化技术来获得有意义的差异化功能。比如，第 1 秒用力呼气量 (FEV1) 可以从肺活量计提供的呼气量 - 时间曲线

### 病人的偏好

不同的病人群体对于自我管理疾病的移动系统的需求和偏好可能有很大不同，这就使定制化的交流方式显得非常重。如图 3a 所示，一位经常接触智能手机应用的用户可能不会介意被手机的警告信息打扰，而另一位用户就可能讨厌这种打扰。在图 3b 中，调查问卷的界面每次只显示一个问题，这符合大多数病人的偏好。图中的调查问卷模块是用于慢性阻塞性肺病监测的移动医疗系统的一部分，但只要换换问题，该模块也可以方便地用于其他移动医疗系统。

按偏好设定关于指标检测、服药或填写调查问卷的提醒天数和次数，让病人获得了更大的灵活性，并能促进病人坚持使用移动设备中介，而不是将设备关掉，关掉设备可能导致检测和用药出现错误。

### 自适应界面

病人必须能够方便直观地和移动中介进行交流，所以设备上所显示信息的复杂程度和形式要由病人的受教育水平、年龄和兴趣来决定。界面适应性从简单的设置调整（比如针对老年用户提高音量）到针对特定年龄群体定制整个界面。

慢性阻塞性肺病病人一般更喜欢简单易懂的反馈信息和简洁的屏幕显示，所以对其健康状况进行可视化呈现更为可取。我们对慢性阻塞性肺病监测系统的实验证实了这些偏好。

另一方面，年轻病人对现代技术更加熟悉，因此可以接受更复杂的界面和更多的功能。他们同样喜欢图标，比如笑脸。



图3. 用于监测慢性阻塞性肺病恶化或妊娠并发症的安卓智能手机移动医疗系统界面截图。(a) 两套系统都使用警告模块来提示病人执行护理任务。(b) 用于慢性阻塞性肺病系统的调查问卷，问题设置相对简单，因为慢性阻塞性肺病病人一般年纪较大。而怀孕病人通常比较年轻，喜欢更详细的信息显示。比如(c) 临床数据，(d) 测量数据，(e) 目前的状态和建议，(f) 预后图以及(g) 测量结果分析。根据病人的偏好定制界面是病人接受移动中介的关键

因此，我们允许使用妊娠并发症监测系统的病人选择病情发展的预后图，预测期直到分娩完成。预后图根据内嵌的贝叶斯网络模型的概率和详细的测量结果分析而绘制。图3e到3g是各自的界面显示。总体看来，图3中的界面是将通用显示元素进行定制，以适应特定病人群体的偏好。用不同的定制化方式，可以使移动医疗系统在其他目标病人群体中同样有效。

获取数据的频率也可以根据病人的风险进行调整。低风险病人需要的检查次数更少，可以降低获取数据的频率。如果病人的风险增加，系统会自动提高

检测频率，因此可以更及时地监测到任何健康恶化的情况，并警告病人和医护人员采取适当的行动。

**信**息和通信技术正在塑造着医疗系统的未来，未来的病人能够更多地参与医疗管理的选择。这些技术不仅要提供连续不断的健康监测，还要提供个人化的病人决策支持和疾病管理方面的建议。

现在，在移动设备上运行大型模型变得更加容易，<sup>14</sup>但医疗是个复杂的过程，需要技术之外的解决方案。决策支持必须

具备临床上的可靠性，因为使用移动医疗系统的病人经常不在可控的临床环境中。适当的训练计划能让临床医生更愿意将个人临床实践与移动决策支持技术相适应。这一技术、临床和心理因素之间的联系，更凸显出临床医生、计算机科学家、工程师和病人在系统开发的早期进行多学科合作的重要性。只有这样，原型系统才能转化成为实用的日常产品。C

## 致谢

我们在此感谢匿名审稿人对本文富有建设性的评论。本文由荷兰健康研究

## 关于作者

**玛丽娜·梅利科娃**是荷兰应用科学组织嵌入式系统创新研究部门的研究人员，她的研究领域包括知识表达、决策支持系统和智能数据分析等。维利科娃在蒂尔堡大学获运筹学博士学位。她是 IEEE 和美国计算机协会会员。联系方式：[marina.velikova@tno.nl](mailto:marina.velikova@tno.nl)。

**彼得·J·F·卢卡斯**是荷兰内梅亨大学计算和信息科学研究所的负责人，荷兰莱顿大学（Leiden University）莱顿先进计算科学研究所教授。他的研究领域包括概率逻辑、概率图模型、决策支持系统和移动医疗解决方案等。卢卡斯在莱顿大学获医学博士学位，在阿姆斯特丹自由大学获得数学与计算科学博士学位。联系方式：[peterl@cs.ru.nl](mailto:peterl@cs.ru.nl)。

**马尔藤·范德海登**是荷兰内梅亨大学计算和信息科学研究所的博士后研究员，他的研究领域包括概率图模型和移动医疗解决方案等。范德海登在荷兰内梅亨大学获人工智能博士学位。联系方式：[m.vanderheijden@cs.ru.nl](mailto:m.vanderheijden@cs.ru.nl)。

与开发组织、STW 技术基金会以及 IT 工程基金会支持的研究成果基础上写成。

## 参考文献

1. Global Observatory for eHealth (GOe), M-Health: New Horizons for Health through Mobile Technologies, tech. report, Worldwide Health Org., 2011.
2. S. Kumar et al., “Mobile Health: Revolutionizing Healthcare through Transdisciplinary Research,” Computer, vol. 46, no. 1, 2013, pp. 28–35.
3. H. Viswanathan, B. Chen, and D. Pompili, “Research Challenges in Computation, Communication, and Context Awareness for Ubiquitous Healthcare,” IEEE Comm., vol. 50, no. 5, 2012, pp. 92–99.
4. A. Triantafyllidis et al., “A Pervasive Health System Integrating Patient Monitoring, Status Logging, and Social Sharing,” IEEE J. Biomedical and Health Informatics, vol. 17, no. 1, 2013, pp. 30–37.
5. M. van der Heijden et al., “An Autonomous Mobile System for the Management of COPD,” J. Biomedical Informatics, vol. 46, no. 3, 2013, pp. 458–469.
6. M. Velikova et al., “Exploiting Causal Functional Relationships in Bayesian Network Modeling for Personalized Healthcare,” Int'l J. Approximate Reasoning, vol. 55, no. 1, 2014, pp. 59–73.
7. Continua Health Alliance, The Next Generation of Healthcare: Personal Connected Healthcare, tech. report, 2010; [www.continuaalliance.org/sites/default/files/continua-alliance-white-paper-next-generation-of-healthcare.pdf](http://www.continuaalliance.org/sites/default/files/continua-alliance-white-paper-next-generation-of-healthcare.pdf)
8. E. Larson et al., “SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone,” Proc. →14th ACM Int'l Conf. Ubiquitous Computing (Ubicomp '12), 2012, pp. 280–289.
9. C. Scully et al., “Physiological Parameter Monitoring from Optical Recordings with a Mobile Phone,” IEEE Trans. Biomedical Eng., vol. 59, no. 2, 2012, pp. 303–306.
10. E.S. Berner et al., “Performance of Four Computer-Based Diagnostic Systems,” New England J. Medicine, vol. 330, no. 25, 1994, pp. 1792–1796.
11. P. Szolovits, “Uncertainty and Decisions in Medical Informatics,” Methods Information in Medicine, vol. 34, nos. 1–2, 1995, pp. 111–121.
12. J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.
13. D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
14. S. Evers and P.J.F. Lucas, “A Framework for Development, Teaching, and Deployment of Inference Algorithms,” Proc. →th European Workshop Probabilistic Graphical Models (PGM '12), 2012, pp. 99–106.

default/



IEEE computer society

# ROCK STARS OF CYBER SECURITY

Win the New Cybersecurity War with  
the New Rock Stars of Cybersecurity

Cybercrime is no longer a matter of credit card breaches. Cybercriminals are now trying to take down countries as well as top companies. Keep your organization safe. Come to the premier, one-day, high-level event designed to give real, actionable solutions to these cybersecurity threats.

Learn from and collaborate with the experts—



CHRIS CALVERT  
Global Director, HP  
Enterprise Solutions Products



MARCUS H. SACHS  
VP, National Security Policy  
Verizon



Dr. SPENCER SOOHOO  
CSO/Director, Scientific Computing  
Cedars-Sinai Medical Center

27 October 2015  
The Fourth Street Summit Center  
San Jose, CA

REGISTER NOW

Early Discount Pricing Now Available!

[computer.org/  
cyber2015](http://computer.org/cyber2015)



# 实现医用级别的 实时移动医疗服务

姜庚泰 ( Kyungtae Kang ) , 韩国汉阳大学 ( Hanyang University )

王启新, 香港理工大学

许峻范 ( Junbeom Hur ) , 韩国中央大学 ( Chung-Ang University )

朴景勋 ( Kyung-Joon Park ) , 韩国大邱庆北科学技术学院 ( Daegu Gyeonbuk Institute of Science and Technology )

夏雷 ( Lui Sha ) , 美国伊利诺伊大学香槟分校 ( UIUC )

一项心电图无线监测的案例研究表明,

当前的 CDMA2000 蜂窝通信技术在医疗遥测领域拥有可观潜力。

相关网络协议栈经改良后, 可实现最大的数据完整性和最短的服务延迟。

# 无

线通信正在迅速成为现代医疗的核心组成部分, 让病人能够享有更高的移动性。病人从有线医疗设备和医疗体系内部的官僚工作作风中获得解放, 由此带来的种种令人叹服的好处, 正是移动医疗 ( m-healthcare ) 系统及应用近来呈爆发式增长的关键动力所在。<sup>1-3</sup> 由于移动医疗重度依赖通过无线网与临床后端系统连接的各类移动医疗设备之间的协作交互, 该领域所面临的关键挑战便是如何达到医用级别的服务质量 ( QoS ) ——一定水平的传输速度、可靠性、隐私性和安全性, 用以提供实时、保密且精确的服务。一套能够提供医用级 QoS 的无线系统, 必须具备足够大的覆盖能力, 足够低的错误率, 以及足够短的服务等待时间上限。

有若干拟想系统均同时支持采用 IEEE 802.11 标准无线局域网和蜂窝通信技术作为医疗遥测应用的平台。<sup>4-6</sup> 无线局域网成本效益高, 是许多医院的首选。蜂窝网络本身成本更高, 但是能够提供面积更广的覆盖能力, 此类基础设施在高移动性所需的支持方面久经考验, 采用授权频段传输也能避免来自其他网络的干扰。此外, 基于既有蜂窝网络基础设施实现的系统, 能够极大地降低成本, 因而在一定程度上抵消了蜂窝网络平台的这一主要缺点。这些特性令蜂窝技术成为了医疗遥测及其他移动医疗应用的一项有力候选技术。<sup>4,5</sup>

为了探讨如何挑选蜂窝网络平台的实际问题, 我们评估了几种重要医疗应用的 QoS 需求, 然后基于 CDMA2000 1xEV-DO<sup>7</sup>

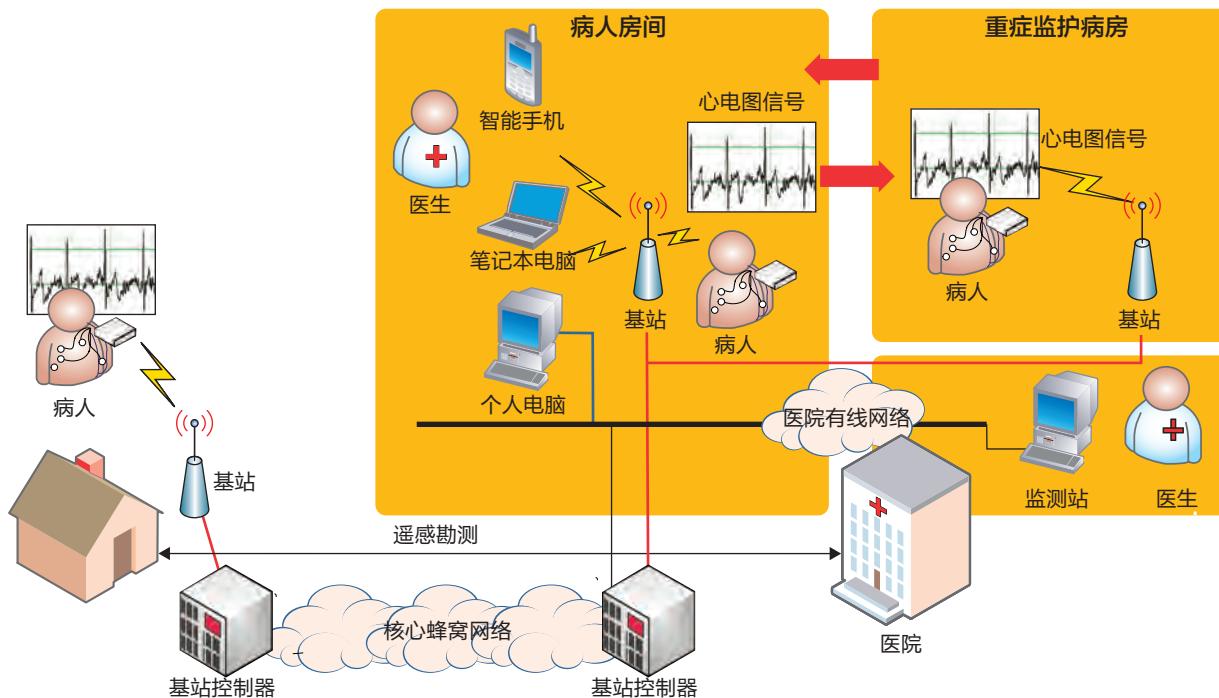
蜂窝网络技术, 开发出一种无线系统架构, 采用码分多址 ( CDMA ) 和时分多址 ( TDMA ) 复用技术, 达到数据吞吐量的最大化。EV-DO ( 数据传输优化 ) 是一种 CDMA2000 1x 演进版本的简写, 代表着一种数据通过无线电信号进行无线传输 ( 主要是用于接入宽带互联网 ) 的通信技术标准。

为了测量我们这种架构的有效性, 并且评估相应的重点 QoS 指标, 我们在经过我们改良的网络协议栈的所有层中, 模拟了心电图 ( ECG ) 的持续性无线监测。在这一点上, 我们的做法与近期其他基于蜂窝网络的移动医疗系统 QoS 的研究有所差别, 后者更多地集中在高层系统设计, 而非网络协议上。<sup>2,6</sup>

为了测试我们的拟想架构, 我们选择了一款推送 QoS 信封的应用。一如其名所示, 心电图的持续性实时监测需要持续不断的实时数据传输。每小时的丢包率最多只能有几秒钟; 错包率必须低于 0.1%, 服务等待时间必须在 2 秒以内。此外, 安全性和隐私性也必须保持在可接受的水平。<sup>8,9</sup>

虽然心电图持续性监测需要的数据速率比许多移动医疗应用都要更高一些,<sup>4</sup> 但它对数据可靠传输和较短服务等待的需要, 与其他那些应用是完全一致的,<sup>8,9</sup> 因为数据丢失或投递滞后显然有损于任何移动医疗应用的效益。故此, 数据包投递率、安全性、服务等待时间和抖动均为至关重要的 QoS 参数, 不仅是对无线心电图持续性监测而言, 也是对全体移动医疗应用而言。

我们的研究并不限于特定应用, 结果显示, 在该领域实现蜂



**图 1.** 基于蜂窝网络技术的心电图持续性无线监测服务。近程无线通信技术，譬如蓝牙，可从体表传感器接收信号，这些传感器专门用于测量贴放在病人体表处的电极之间的电压差。信号接着被输送至病人的移动设备，该设备再通过中间基站控制器将数据传送给基站。然后再由基站将数据发送至病人房间内或重症监护病房内的医院监控站

窝网络时，需要仔细权衡各方利弊。一边是更佳的病人移动性和更低的部署成本；一边是在传输延迟有所改善的同时，可靠性及安全性却会遭到潜在的弱化。要了解最优的制衡结果，需要进行谨慎的分析。

## 心电图持续性监测的要求

在一套心电图的持续性无线监测系统中，病人会在附近如常走动，临床医生则会即时获取他们的心电图数据。就这一层面而言，持续性监测不同于那些收集并存储数据以供稍后下载的系统，也不同于那些仅在特定事件（例如心律不齐或心搏骤停）发生时才会传输数据的系统。

### 通过蜂窝网络进行监测

图 1 所示为一种利用蜂窝网络系统

进行心电图持续性无线监测的可能性场景。

该系统必须设法对心电图模拟信号进行取样和数字化，并分批转化为数据包。具体选用的取样频率和数字化方法，将会决定传输过程中的流量特性。数据速率的大小，则取决于传感器的数量，还有取样频率。如果有  $L$  只传感器对病人心搏的电子信号进行取样，而系统对每只传感器的输出信号进行数字化的频率为  $r$  个样本 / 秒，精度为  $S$  位，则最终得到的数据速率为  $(LrS)$  位 / 秒。因此，若有 12 只传感器以 16 位的精度每秒提供 500 个样本信号，最后便可获得 96Kbps 的数据速率——这正是蜂窝网络心电图监测应用的常用数值。该范例演示了在设计蜂窝基础设施时，目标数据速率对心电图监测应用的重要性。

出于对心电图持续性无线监测 QoS 的严格要求，需要对网络协议栈进行若干改良，尤其是在数据链路层和物理层。

**数据链路层。** 链路控制可通过三项主要功能，提供可靠安全的数据通信。

安全功能，可为设备和网络提供认证，确保数据在传输过程中不会遭到修改，并对传输数据进行加密，保证隐私性。

逻辑链路控制，可为无线信道内部的突发差错进程提供补偿。差错控制可基于重传进行，系统将只重新传输差错部分，并且很容易适配不断变化的信道状况。然而，数据递交时间却无从预测，这在移动医疗应用中是无法接受的。显而易见的替代方案是前向纠错（FEC），它能够在维持吞吐率的同时控制延迟。

数据链路层的第三项功能是介质访问控制（MAC），用于决定设备何时可以访问通信介质。CDMA 和 TDMA 两种

### 提供医用级 QoS

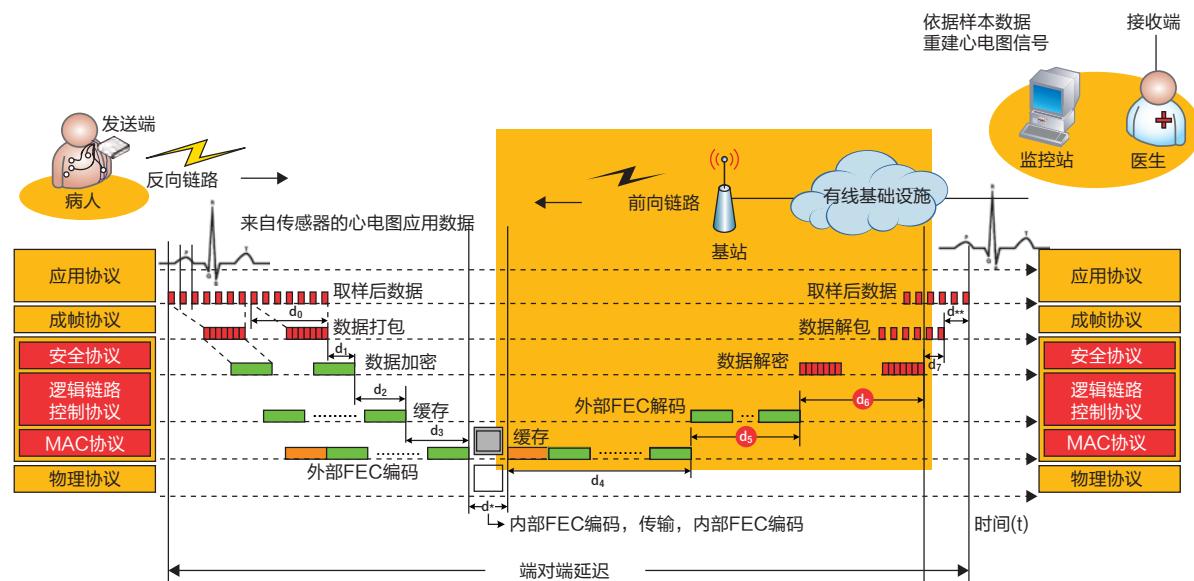


图2. 改良后的协议栈和延迟来源。该协议栈依据在心电图持续性监测中采用的CDMA2000 1xEV-DO(码分多址联接方式, 数据传输优化)蜂窝网络技术而设计。延迟来源( $d_n$ )散布在各层的不同功能中。黑色圆框中的延迟视无线信道的状况而异。FEC: 前向纠错; MAC: 介质访问控制

技术,总有一种能够满足移动医疗应用的计时要求,与IEEE 802.11网络目前所采用的随机访问或基于争用的方法相比,可让延迟变得更容易预测。

**物理层。**在物理层完成的信道编码和数字信号调制,目的是优化吞吐量和可靠性。信道编码在数据上加入冗余位,用于处理因噪音或干扰而出现缺损的数据位。诸如纠错编码、卷积编码等这类FEC技术,一般均可提供这种冗余处理。当物理层数据包的长度达到数百位甚或更长时,纠错编码技术的效果会更佳。

传输比特率和可靠性之间的此消彼长,牵动着对调制类型和FEC码率的选择:越是稳健的调制类型,对干扰水平的容许度更高,但传输速率却会更低。更高的FEC码率将会加大冗余度,让系统能够容许更高水平的干扰,这会提高数据传输的可靠性,但也会降低有效的比特率。一般而言,移动医疗应用更适合注重可靠传输多于高速传输的调制及编码机制,

但如果系统架构师清楚地知道需要达到怎样的性能平衡,也可采用调制及编码的自动适配。

## 采用CDMA2000技术实现的无线心电图

获知具体的QoS要求后,我们便开始着手研究如何才能让相关协议栈适用于CDMA2000 1xEV-DO蜂窝网络系统下的心电图持续监测。

### 协议层

图2所示为我们拟想的协议栈及各层之间的交互示意图。这是在CDMA2000多播服务标准空中接口规格的基础上设计而成的。<sup>10</sup>成帧协议用于将更高层的数据打包成帧,该协议可帮助确定更高层数据的分界。然而,较低层的协议却需要进行大幅度的修改。

**安全层与数据链路层。**安全层的数

据加密使用的是高级加密标准(AES),用一个128位的密钥对128位的数据块进行加密。<sup>10</sup>数据链路层的外部FEC编码与物理层的内部FEC(纠错)编码相结合,形成一个有效的乘积码。我们的协议栈使用的是理德·所罗门(RS)算法的外部编码,理由在于该算法在低差错率上的卓越表现——对心电图持续性监测而言,这是理想之选。

我们将RS码设定为(N,K),即编码器获取K个信息码元,并在后面加上(N-K)个监督码元,从而得到一个码长为N的码字。一个解码器最多可纠正t个符元错误(位置不详),或2t个删除错误(位置已知的错误),其中 $2t = N - K$ 。删除符元的位置通常可在数字通信系统中查得,因为解调器普遍都会标示出可能包含错误的传入符元。

由于FEC码专为处理随机差错而设计,无法应对突发差错,于是我们加入了交织处理,<sup>11</sup>这个简单的处理能够依次散播突发差错,好让FEC对其进行纠错,

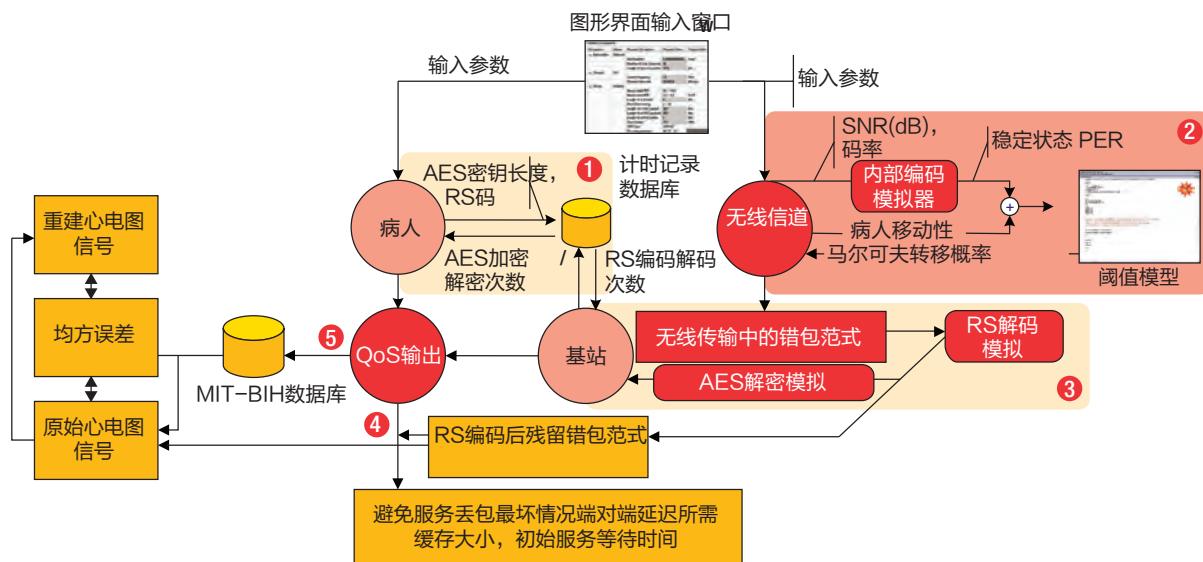


图3. 心电图持续性无线监测系统的模拟示意图。(1) 系统在理德一所罗门(RS)编码和高级加密标准(AES)加密解密过程中得出延迟时长;然后(2)在使用阈值模式进行无线传输的过程中形成错包。然后由基站(3)模拟RS解码和AES解密过程。最后,系统(4)分析服务质量(QoS)指标,并用所得结果(5)估算出均方误差;SNR:信噪比

如此一来, RS 码也可处理突发差错了。

为了处理发送端的突发差错, 系统会将安全层数据插入缓冲存储器, 或者按照从左到右、自上而下的顺序, 填充至差错控制编码块(ECB)即一个由 N 行和 M 列构成的表中, 然后依次对每一列应用 RS 编码算法。与之相反的解码及解包过程, 则在接收端进行。交织算法能够纠正的突发差错长度, 取决于 ECB 的宽度。M 值越大, 能够纠正的突发差错长度就越长, 不过代价是所需的缓冲区会更大, 等待时间也会更久。<sup>11</sup>

**介质访问控制层。**CDMA2000 的 MAC 协议同时指定了前向业务信道和反向业务信道两者的数据传输方式。反向链路利用 CDMA 技术和沃尔什编码来传输数据, 采用长码伪噪声序列识别移动设备。前向链路同时利用 CDMA 和 TDMA 技术来传输数据。

**物理层。**CDMA2000 1xEV-DO 修订

版 A 系统可支持的反向链路数据速率为 4.8 ~ 1,843.2Kbps。具体速率视有效负载、调制类型和纠错编码中加入的冗余度而定。

### 延迟来源

如图 2 所示, 运行在 CDMA2000 网络上的 ECG 应用有多处端对端延迟来源。在每帧长度为 1,000 位, ECG 数据速率为 96Kbps 的情况下, 填充一帧所产生的积累延迟( $d_0$ )为 10.4 毫秒。加密过程会带来一段附加延迟( $d_1$ )。根据该帧长度、RS 码为(16,12)、ECB 的每一行中最多可填充两个安全层数据包( $M = 2$ )来计算, 写入 ECB 的过程中总共会发生 250 毫秒的延迟( $12 \times 2 \times 1,000/96$ )( $d_2$ )。

将安全层数据包逐行输入 ECB, 然后逐列执行外部 FEC 编码, 又会产生一段附加延迟( $d_3$ )。在物理层的硬件(包括内部 FEC 编码器和解码器)中, 自然还会发生进一步的编码和传播延迟( $d^*$ ),

不过在我们的分析中, 这些延迟小到可以忽略不计。

基站执行内部 FEC 解码, 然后将数据包缓存至 ECB 内, ECB 的数据填充速率是固定的, 因为 CDMA2000 在访问介质时遵循的是 TDMA 协议。因此, 缓存延迟( $d_4$ )是一个常数。一俟 ECB 写满, 基站便会执行外部 FEC 解码, 解密安全层数据包, 监测站再将其解包, 获取原始的 ECG 数据。这些动作会产生 3 段附加延迟( $d_5$ 、 $d_6$  和  $d_7$ )。(我们不考虑应用处理延迟( $d^{**}$ )的影响, 因为该延迟与无线系统无关。)

所有这些延迟都是固定不变的, 只有 RS 解码过程中产生的延迟( $d_5$ )和 AES 解密过程中产生的延迟( $d_6$ )例外, 这两个延迟具体将视物理信道的状况而异。利用软件实现这两个环节, 能够大幅度地减少这种延迟抖动。采用这种实现方式时, 系统将需要一处缓存区来吸收这种延迟抖动, 这种抖动的最大振幅为

$$2 \times \left( \overline{d_5 + d_6} - \underline{d_5 + d_6} \right)$$

其中

$$\overline{d_5 + d_6}$$

为 ECB 的 RS 解码和 AEC 解密的最坏情况时间；而

$$\underline{d_5 + d_6}$$

则是相应的最佳情况时间。抖动缓存可将这种最大程度的抖动转变成同等大小的固定延迟 ( $d_7$ )。

## QoS 分析

优化心电图无线监测网络，需要考虑可靠性与等待时间在同一层内和不同层之间的平衡。为了分析这种冲突的具体情况，我们用 Java 做了一次心电图持续性监测应用的跨层模拟。图 3 所示为此次模拟的整个过程，其间我们曾不断变换信道参数和网络参数，以便获得相应的 QoS 指标。

此次模拟使用了 RS 删码器和解码器的软件实现 (<http://rscode.sourceforge.net>)，以及格莱德曼的 AES 算法参考实现 ([www.gladman.me.uk](http://www.gladman.me.uk))。我们假定病人的移动设备采用的是低功耗的 ARM9E 处理器内核，但也假定基站配备的有可能会是功耗更为强劲的 ARM11 内核。之所以选择 ARM 处理器，是因为许多适配 GSM 和 CDMA2000 网络的 3G 手机所使用的都是此类处理器。

我们还选用了支持 ARM 处理器的 IAR 集成开发环境 ([www.iar.com](http://www.iar.com))，这

表 1. 无线心电图监测系统参数一览

参数	值
网络	
内部纠错码率 ( $r$ )	1/4
物理层数据包长度	1,024 位
样本外部 RS 码 ( $N, K$ )	(16,12)
RS 删码中的符元位数 ( $s$ )	8 位
编码块交织层级 (M)	1–8
安全层数据包长度	1,000 位
MAC 层数据包长度	1,002 位
AES 加密块单位长度	128 位
加密密钥长度	128 位
物理信道	
载波频率	1.8 GHz
调制类型	相移键控
信道数据速率参考值	153.6 Kbps
信道信噪比最大值	-2.77 dB
预估无线信道移动速率	2–4K mph
无线心电图	
传感器数量	12
每只心电图传感器每秒取样频率	500 Hz
取样长度	16 位
数据速率	96 Kbps
医疗设备	
病人移动设备	ARM9TDMI，工作频率 250 MHz
基站（现代配备）	ARM11，工作频率 400 MHz

MAC：介质访问控制；RS：理德—所罗门；AES：高级加密标准

款工具让我们能够分析出安全层中 AES 加密解密的执行时间，以及逻辑链路层中 RS 编码解码的执行时间。

## 模拟参数

表 1 列出了我们在此次模拟中的各项参数，数据出自 CDMA2000 1xEV-DO 修订版 A 的标准配置。由于移动设备的射频功率有限，需要选择受信道恶劣条件影响较少的码率和调制技术。因此，我们选择的反向信道数据速率为 153.6Kbps，RS 码为 (16,12)，从而为编码增益和处理延迟提供了一个适宜的平衡点。<sup>10</sup>

## 模拟过程

此次模拟首先得出了病人设备（发

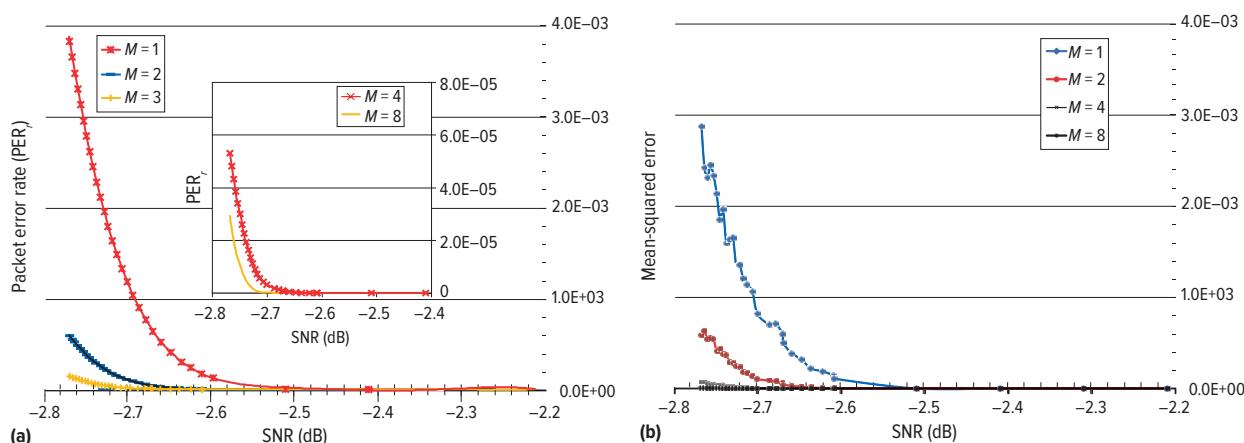
送端）中 AES 加密所产生的延迟 ( $d_1$ ) 和 RS 编码所产生的延迟 ( $d_2$ )。接着利用编码调制库 ([www.iterativesolutions.com](http://www.iterativesolutions.com)) 模拟物理层的内部纠错编码，基于给定信噪比，又得出了该信道的错包率 (PER)。信道错包率以及病人的移动速度，为模拟这种无线信道模型中数据包突发差错的发生概率提供了基础。<sup>12</sup>

为物理层中的错包过程建立起模型后，便可实现对基站活动的模拟。模拟逻辑链路层中的 RS 解码过程和安全层中的 AES 解密过程，可得出每块 ECB 残留误差的范式及数量，同时还有 RS 解码和 AES 解密过程中各自产生的延迟 ( $d_4$  和  $d_5$ )。

以上结果被递交至输出控制模块，模块将残留误差注入 MIT-BIH 心律失常

表 2. 模拟所得服务质量指标：延迟与等待时间（毫秒）和抖动缓存和 ECB 大小（位）。

$M$	$\overline{PER}_r$	$d_0$	$d_1$	$d_2, d_4$	$d_3$	$\overline{d_5 + d_6}$	$d_5 + d_6$	$d_7$	抖动缓存大小	最坏情况下的服务等待时间	ECB 大小
1	3.3E-3	10.4	7.04	125	10.9	44.91	27.77	34.29	3,291	347.14	16K
2	4.8E-4	10.4	7.04	250	21.8	89.82	55.42	68.81	6,606	687.48	32K
3	1.4E-4	10.4	7.04	375	32.7	131.99	83.10	97.78	9,386	1,019.51	48K
4	4.1E-5	10.4	7.04	500	43.6	174.15	110.81	126.68	12,161	1,351.47	64K
5	3.8E-5	10.4	7.04	625	54.5	214.93	138.55	152.78	14,666	1,679.25	80K
6	2.7E-5	10.4	7.04	750	65.4	257.10	166.25	181.70	17,443	2,011.24	96K
7	2.5E-5	10.4	7.04	875	76.3	297.88	193.93	207.92	19,960	2,339.14	112K
8	2.4E-5	10.4	7.04	1,000	87.2	337.30	221.64	231.32	22,207	2,662.86	128K

图 4. 心电图信号重建的均方误差平均值。随着  $M$  值的加大，差错率仅出现最低限度的下降，然而等待时间却会增加，也就是说，当  $M$  大于 4 时，性能增益十分有限

数据库，<sup>13</sup> 数据库内含 48 条各半小时长的不同心电图数据记录。接着由控制模块估算出以下数据：均方误差，一种用于量化远程监测站上重建的心电图信号与病人设备发送的原始信号之间误差的数据；所需 ECB 和抖动缓存的大小；服务等待时间。

## 模拟结果

我们的模拟结果揭示出了一些有趣的规律。举例来说，模拟结果表明，AES 解密过程处理单个安全层数据包所产生

的延迟 ( $d_1$ ) 会是 7.04 毫秒，而打包心电图数据的过程则会产生 10.4 毫秒的延迟 ( $d_0$ )。因此，到系统对一个帧进行打包时，上一帧的加密过程将会已经完成，这意味着安全层中的 AES 加密过程只会让缓存产生 7.04 毫秒的延迟。对于特定大小的 ECB 而言，两次缓存延迟 ( $d_2$  和  $d_4$ ) 的数值只会取决于心电图数据的取样率。

## 数据包差错

表 2 第二列所示为数据链路层的残

留错包率  
 $(\overline{PER}_r)$

之上限，分别对应编码块交织参数  $M$  的不同值。该表还列出了 RS 解码和 AES 解密的最坏情况执行时间和最佳情况执行时间，以及最大生成抖动导致的相应延迟。最后一列所示为用于处理这一抖动的缓存大小和相应的服务等待时间。

我们还发现了一个与  $M$  值、差错率和等待时间相关的有趣规律。虽然加大  $M$  值能够扩大 FEC 可以处理的突发差错

表 3. 加密一个 128 位数据块和编码一个 RS 码字的能耗

加密		CPU	MEM	合计
		35.412	22.925	58.337
加密	(16,12)	15.130	9.371	24.501
	(16,13)	13.190	8.242	21.432
	(16,14)	11.090	7.143	18.233
	(32,24)	43.276	24.235	67.511
	(32,26)	36.768	20.911	57.679
	(32,28)	28.226	16.314	44.54

CPU：中央处理单元；MEM：主存储器

的时间范围，但是这种变通的代价将是存储器延迟和缓存延迟。然而，我们的模拟却显示，纠错表现在 M 增加到一定阈值后便会趋于饱和，直到所有的差错都会得到纠正。因此，如图 4 所示，加大 M 值，只会在最低程度上进一步降低差错率，而等待时间却会不断增加，如表 2 所示。

在我们的心电图应用中，将 M 值设为 4 或 5，便可得到 99.99% 的可靠度，均方误差平均值在 7.7E-7 以内，等待时间在 2 秒以内。我们假定医院有足够的基站，可将信噪比保持在 -2.77 dB 以上。

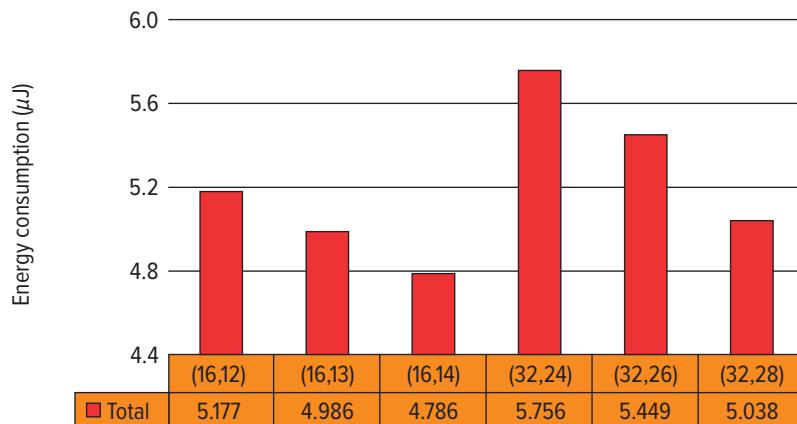
## 能源效率

用于心电图监测的无线技术，不仅必须按时、精确地递交心电图数据，还必须妥善管理一部电量有限的设备中的电能。在我们的拟想无线系统中，数据链路层的 AES 加密和 RS 编码过程的能耗是最高的。

我们使用了 XEEMU 的一款模拟器<sup>14</sup> 来测算 AES 密钥和 RS 编码器软件实现参考的平均能耗。XEEMU 可为使用的 ARM9E 指令集的 ADI 80200EVB XS-scale 模拟板提供精确的能耗数据。

我们假定：用于 AES 加密和 RS 编码（病人设备）的目标处理器内核，操作频率为 250MHz；指令和数据分别使用不同的高速缓存，容量各为 32KBs；同步动态存储器为 128MB 的 Micron SDRAM，时钟速度为 100MHz ([www.micron.com](http://www.micron.com))。

表 3 所示为加密 128 位数据块和采



用常用 RS 码编码码字时所需的能耗。图 5 所示为数据链路层处理 1 字节数据有效负载时所用掉的能耗。

表 3 显示，数据链路层处理 1 字节数据有效负载时，所用能耗更多，因为要提高 RS 码的纠错能力，就要增加监督码元的位数。这一结果在我们的意料之中，因为计算这些校验子和删除求值多项式的复杂度，会随监督信息的总量增加而上升。举例来说，RS 码为 (16,12) 的 RS 编码过程要比 RS 码为 (16,14) 的相应过程，多用掉 36.4% 的能耗。由此可推知，心电图应用需要在保证达到所要求的 QoS

水平的同时，设定一个能够显著节省能源的 RS 码。其中的平衡点取决于移动医疗设备可能遭遇的信道状态，以及需要实现的续航时长。

我们示范的这套基于蜂窝网络的心电图无线监测系统的模拟结果表明，总体而言，当前的 CDMA2000 蜂窝网络技术在实时医疗遥测和移动医疗应用领域拥有可观潜力。因此，我们期待会看到更多的升级版 4G 技术支持个人化移动医疗服务方兴未艾的增长势头。C

## 关于作者

**姜庚泰 (Kyungtae Kang)** 是韩国汉阳大学计算机科学与工程系的助理教授。

他的研究范畴包括操作系统、移动系统、分布式系统和信息物理系统。姜教授拥有韩国首尔国立大学颁发的电气工程与计算机科学博士学位。他是电气和电子工程师协会 (IEEE) 和美国计算机协会 (ACM) 的会员。联系方式: ktkang@hanyang.ac.kr。

**王启新** 是香港理工大学计算系的助理教授。他的研究范畴包括信息物理系统、实时系统、嵌入式系统、实时网络、无线技术，以及这些技术在工业控制、医药和辅助生活领域的应用。王教授拥有美国伊利诺伊大学香槟分校颁发的计算机科学博士学位。他是电气和电子工程师协会 (IEEE) 和美国计算机协会 (ACM) 的会员。联络方式为 csqwang@comp.polyu.edu.hk。

**许峻范 (Junbeom Hur)** 是韩国中央大学计算机科学与工程系的助理教授。他的研究范畴包括信息安全、移动计算和无线网络安全。许教授拥有韩国先进科学技术学院颁发的计算机科学博士学位。他是电气和电子工程师协会 (IEEE) 会员。联络方式为 jbhur@cau.ac.kr。

**朴景勋 (Kyoung-Joon Park)** 是韩国大邱庆北科学技术学院信息与通信工程系的助理教授。他的研究范畴包括信息物理系统的调制与分析，和无线医疗保健系统的医用级协议设计。朴教授拥有韩国首尔国立大学颁发的电气工程与计算机科学博士学位。他是电气和电子工程师协会 (IEEE) 会员。联络方式为 kjp@dgist.ac.kr。

**夏雷 (Lui Sha)** 是美国伊利诺伊大学香槟分校的唐纳德·吉利斯讲席教授 (Donald B. Gillies Chair Professor)。他的研究范畴包括信息物理系统、形式化降复杂度架构范式 (formalized reduced complexity architecture patterns)、分布式实时容错计算系统和动态实时架构。他拥有美国卡耐基梅隆大学颁发的电气与计算机工程博士学位。他是电气和电子工程师协会 (IEEE) 和美国计算机协会 (ACM) 的会士。联络方式为 lrs@ illinois.edu。

## 鸣谢

本研究有部分得到了韩国资讯通信技术暨未来规划部的支持，隶属于信息技术研究中心支持课题（课题编号 NIPA-

2014-H0301-14-1044），由韩国国家信息通讯产业促进局负责督导；部分通过韩国国家研究基金会 (NRF) 得到了自然科学研究项目的支持，由韩国资讯通信技术

暨未来规划部进行资助（课题编号 NRF-2013R1A1A1059188）；还有部分得到了香港研究资助局 (RGC) 青年学者计划 (ECS) 支持，课题编号 PolyU 5328-12E、Hong Kong PolyU A-PJ80、A-PK46 和 A-PL82。

## 参考文献

1. R.S.H. Istepanian, E. Jovanov, and Y.T. Zhang, "Beyond Seamless Mobility and Global Wireless Healthcare Connectivity," *IEEE Trans. Information Technology in Biomedicine*, vol. 8, no. 4, 2004, pp. 405–414.
2. R.S.H. Istepanian, S.P. Costantinos, and S. Laxminarayan, "Ubiquitous M-Health Systems and the Convergence Towards 4G Mobile Technologies," *M-Health: Emerging Mobile Health Systems*, R.S.H. Istepanian, S. Laxminarayan, and C.S. Pattichis, eds., Springer, 2006, pp. 3–14.
3. Y.M. Fang, "Wireless Healthcare: Technologies for Bettering Our Life," *IEEE Wireless Comm.*, vol. 17, no. 1, 2010, pp. 2–3.
4. S.D. Baker and D.H. Hoglund, "Medical-Grade, Mission-Critical Wireless Networks," *IEEE Eng. in Medicine and Biology*, vol. 27, no. 2, 2008, pp. 86–95.
5. Q. Wang et al., "Building Robust

- with the DSSS-CDMA Cell Phone Network Paradigm," IEEE Trans. Mobile Computing, vol. 6, no. 6, 2007, pp. 706–719.
6. D. Vouyioukas, I. Maglogiannis, and D. Komnafos, "Emergency M-Health Services Through High-Speed 3G Systems: Simulation and Performance Evaluation," Simulation, vol. 83, no. 4, 2007, pp. 329–345.
7. N. Bhushan et al., "CDMA2000 1xEV-DO Revision A: A Physical Layer and MAC Layer Overview," IEEE Comm., vol. 44, no. 2, 2006, pp. 37–49.
8. IEEE Std. 11073: Health Informatics—PoC Medical Device Communication, Part 00101: Guidelines for the Use of
- RF Wireless Technology, Dec. 2008.
9. L. Skorin-Kapov and M. Matijasevic, "Analysis of QoS Requirements for E-Health Services and Mapping to Evolved Packet System QoS Classes," Int'l J. Telemedicine and Applications, 2010; [www.hindawi.com/journals/ijta/2010/628086](http://www.hindawi.com/journals/ijta/2010/628086).
10. P. Agashe, R. Rezaifar, and P. Bender, "CDMA2000 High Rate Broadcast Packet Data Air Interface Design," IEEE Comm., vol. 42, no. 2, 2004, pp. 83–89.
11. K. Kang, "Probabilistic Analysis of Data Interleaving for Reed-Solomon Coding in BCMCS," IEEE Trans. Wireless Comm., vol. 7, no. 10, 2008,
- pp. 3878–3888.
12. M. Zorzi, R.R. Rao, and L.B. Milstein, "Error Statistics in Data Transmission over Fading Channels," IEEE Trans. Comm., vol. 46, no. 11, 1998, pp. 1468–1477.
13. G.B. Moody and R.G. Mark, "The Impact of the MIT-BIH Arrhythmia Database," IEEE Eng. in Medicine and Biology, vol. 20, no. 3, 2001, pp. 45–50.
14. Z. Herczeg et al., "Energy Simulation of Embedded XScale Systems with XEEMU," J. Embedded Computing, vol. 3, no. 3, 2009, pp. 209–219.

# IT Professional

TECHNOLOGY SOLUTIONS FOR THE ENTERPRISE

## CALL FOR ARTICLES

*IT Professional* seeks original submissions on technology solutions for the enterprise. Topics include

- emerging technologies,
- cloud computing,
- Web 2.0 and services,
- cybersecurity,
- mobile computing,
- green IT,
- RFID,
- social software,
- data management and mining,
- systems integration,
- communication networks,
- data center operations,
- IT asset management, and
- health information technology.

We welcome articles accompanied by Web-based demos. For more information, see our author guidelines at [www.computer.org/itpro/author.htm](http://www.computer.org/itpro/author.htm).

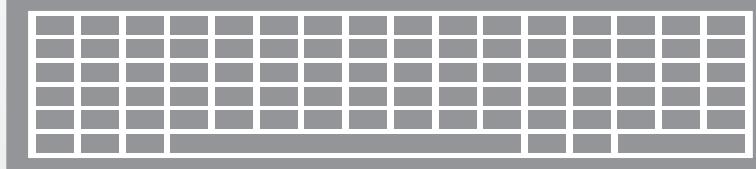
**WWW.COMPUTER.ORG/ITPRO**





学习  
创业 市场 找工作  
职场 行业分析 申请 合作  
竞赛 进修 教育  
科技新闻  
研究综述

内容



# 医疗数据整合 与云端信息学

阿什迪普·巴哈 (Arshdeep Bahga)、维杰·K·马蒂塞蒂 (Vijay K. Madisetti)，佐治亚理工学院 (Georgia Tech)

用于医疗应用的信息整合和信息学框架，能够利用由商用硬件组成的云端大规模分布式批处理基础设施的并行计算能力，为高级医疗应用的开发者提供新的灵活性。

**由**于医疗类应用数据的种类繁多，数据量庞大，所以这类应用的开发者和人口健康研究人员面临的主要挑战就是如何对这些数据进行整合和有效分析。传统的医疗信息技术系统，如电子健康记录 (EHR) 和个人健康记录 (PHR) 系统各自基于相应架构，使用不同的技术和语义标准来表示和存储数据。这些客户端服务器系统取决于本地的硬件、软件和数据存储，每套系统都可以使用不同的语言和数据库技术。上述所有这些特点，使得对来自多个时常冲突的系统的数据进行准确而便捷地整合变得极为困难，而这种整合恰恰是开发高级医疗应用的关键。

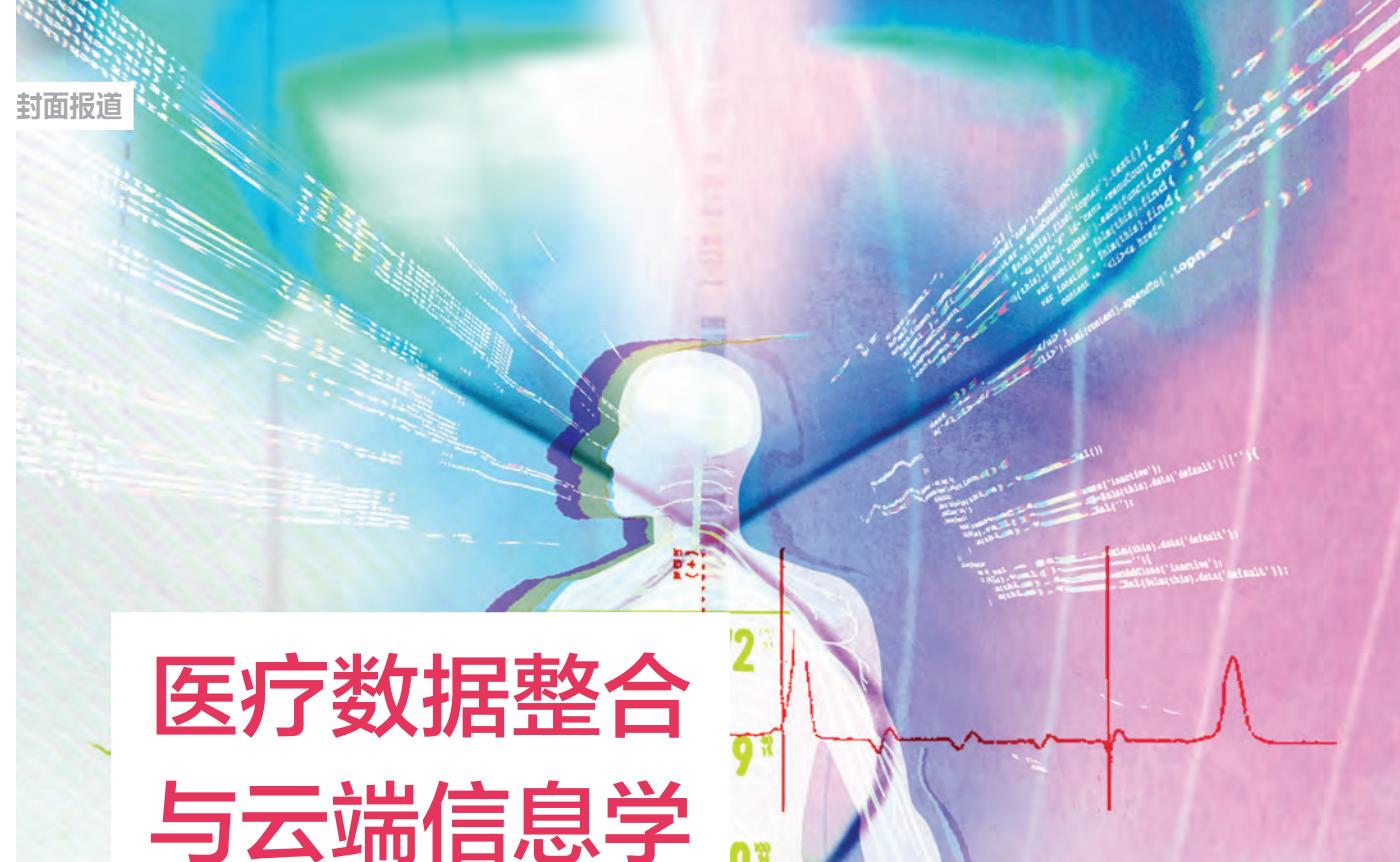
相比之下，基于云的系统允许将数据存储在外部服务器上，<sup>1</sup> 开发者可以方便地访问。但互操作性仍然是个大问题，本文中的插文《电子健康记录系统的互操作性》讨论了这个问题。在早期研究中，<sup>2</sup> 我们解决这一问题是通过在可互操作的电子健康记录的系统设计中创建云端访问入口，并将该系统并入云健康信息系

统技术架构 (CHISTAR)，这是一套可实现语义互操作的原型系统。CHISTAR 类架构在设计方法上的参照模型定义了一套通用的数据结构，其原型模式则定义了临床数据的属性。CHISTAR 能够实现对健康数据的安全访问，并支持授权、身份管理和认证服务等功能。

我们将这一成果进行扩展，把云端信息的整合和使用可靠的开源云端技术的信息学 (三级) 框架包括进来，以促进对来自不同地理位置的临床数据的收集和分析。其功能包括：

- ▶ 将分布式异构资源整合为通用命名
- ▶ 对储存在云中的医疗数据的访问更加容易
- ▶ 对云中收集的大量医疗数据的分析效率更高
- ▶ 医疗数据存储和终身管理

我们的三级框架的重要优势在于对临床数据整合和分析技术



## 电子健康记录系统的互操作性

的使用，这些技术能够利用在其他领域已经应用的云计算环境产生的效益和经济作用。随着来自分布式异构资源的临床数据量继续呈指数级增长，数据分析越发成为问题所在，并成为更复杂的医疗应用开发的瓶颈。如何对收集自分布式异构健康 IT 系统的临床数据进行整合，将对更有效的医疗应用的诞生发挥更大的作用。对大规模临床数据的分析方式将促进更有效的医疗应用的开发，提高预测的准确性，有助于及时做出决策。

用我们的三级框架与 CHISTAR 中间件相结合，可以对包括病人、医院、理疗师、保险公司在内的一系列利益相关人的医疗数据进行收集、组织和安全交换，数据的格式也包括数据库、结构性和非结构性数据等多种。如此规模的数据交换有助于确保治疗的及时性和准确性，因此在降低医疗成本的同时提升了质量。此外，这一框架有助于开发高级医疗应用，比如对流行病的监测和对不良药物事件的预测等。

流行病监测在诸多高级应用中尤其具有代表性，<sup>3</sup>因为它涉及在指定的人群中，研究与健康相关的状况和事件的分布和决定因素，并将任何发现用于全国监测下的疾病诊断。正因如此，我们选择流行病监测作为支持开发者开发高级医疗应用的一个范例。我们随后的分析显示，对于流行病监测这样的应用，我们的三级框架相对于客户端 - 服务器电子健康记录系统具有几点优势，包括更强的可扩展性，更快的开发速度和更低的成本。

**由**于电子健康记录数据和系统互操作性是高级医疗应用的基础，所以关于如何实现可能的互操作性的研究越来越多。语义互操作性是 OpenEHR ([www.openehr.org](http://www.openehr.org)) 的关注焦点，比如名为 Mirth Connect 的开源整合引擎，就支持多种信息标准和协议，可连接至外部系统和大量数据库来储存信息数据。

一项研究描述了电子健康记录数据对流行病监测的潜力，但目前监测大型区域态势的方法与在简仓中运行的医疗系统并不相容。为了在这种情况下提高互操作性，研究团队针对电子健康记录数据提出了一套肺结核监测算法，以及实时监控的实现和报告系统。

另一项研究则提出了基于电子健康记录的监测方法，来确定冠心病的高危因素，并给出了一个基于电子健康记录的人口水平监测模型。

### 参考文献

1. M.S. Calderwood et al., "Real-Time Surveillance for Tuberculosis Using Electronic Health Record Data from an Ambulatory Practice in Eastern Massachusetts," *Public Health Reports*, vol. 125, 2010, pp. 843–850.
2. J.J. VanWormer, "Methods of Using Electronic Health Records for Population-Level Surveillance of Coronary Heart Disease Risk in the Heart of New Ulm Project," *Diabetes Spectrum*, vol. 23, no. 3, 2010, pp. 161–165.

## 架构总览

如图 1 所示，三级框架是堆栈的一部分，堆栈中包括信息学应用生成器和 CHISTAR 中间件。为了方便用户界面的开发，中间件为安卓、iOS 等移动操作系统和 Windows 桌面应用提供了平台工具。该框架还为数据的整合、访问、分析和储存提供了工具和应用程序编程接口（API）。

### 应用生成工具

顾名思义，信息学应用生成器是用

于构建移动和网络医疗应用的工具套件。由于三层框架和 CHISTAR 中间件负责底端的云基础设施管理、部署配置和数据管理，开发者可以使用信息学应用生成器，而不用考虑云端的配置和维护活动。

### 确保互操作性

CHISTAR 中间件提供一系列表示、应用和信息服务支持高级信息学应用的开发。其安全服务满足《健康保险携带与责任法》(HIPAA) 和《经济和临床医疗信息技术法》(HITECH) 中的关键要求，这两部法律要求管辖范围内的组织（即



**图1** 用于构建医疗应用的技术堆栈，需要从分布式异构电子健康记录（EHR）系统中获取大量的临床数据。信息学应用生成器提供应用开发工具，云健康信息系统技术架构（CHISTAR）中间件可以保证语义的互操作性，信息整合和信息学（三层）框架负责数据的整合、访问、分析和储存

生成、维护、传输、使用、披露受保护的个人健康信息的组织）保护其所收集、维护、传输、使用数据的完整性、机密性和有效性。

CHISTAR的应用服务是独立平台，开发者可以定制工具来支持指定的平台功能，为多种多样的平台和设备进行医疗应用开发。

表1中列出了CHISTAR的安全服务。

## 数据整合

数据整合方法主要因面对层级的不同而变化。在应用层上，整合是通过在域范围的应用上重新应用从个人应用中整合而来的数据。在API层的整合上，应用会开放其编程接口，使其它应用可

以访问其数据。最后，在数据层上，整合将建立一个通用域模型或全局模式，让独立开发的应用能够交换信息。我们的三级框架中并入的正是数据层的整合，在开发上的灵活性比其他方式更高，因为它允许独立开发，并着眼于通用数据交换。

## 框架实现

我们的三级框架由多个层级构成，首先是CHISTAR中的数据整合和交换引擎，构成数据整合层。数据访问层使用一系列开源技术来实现对云端数据的访问。

本框架通过数据存储模型来表示数据结构，通过域模型来表示临床知识，从而实现数据的互操作性。

数据编配层由HBase和Zookeeper组成。HBase是一个非关系型、面向列的分布式数据库，运行于Hadoop分布式文件系统（HDFS）上，是Hadoop的一部分，Hadoop是一套云端分布式批处理基础设施。HBase可以提供对大量零散数据的一种容错储存方式。Zookeeper是一项分布式协作服务，负责维护配置信息、命名，提供分布式同步和集群服务。数据分析层建在Hadoop层之上。

最后，数据存储层由分布式文件系统和存储源文件（如图片）的云存储组成。整合、访问和分析层的实现是最具挑战性的部分。

## 数据整合

如图2所示，数据整合可以对以不

表 1 CHISTAR 中间件的安全服务

服务	基础
身份验证	开放授权 ( OAuth )
授权	专用入口或客户服务器
身份管理	联邦身份
静态数据安全	高级加密标准 ( AES ) 加密
动态数据安全	安全套接层 ( SSL )
密钥管理	密钥储存、循环和加密使用独立密钥
数据完整性保证	信息认证码
审核	全部读写行为的日志记录

同形式（机构化或非结构化）的健康数据进行整合，数据可以存在于不同的数据存储系统中，如关系型数据库管理系统（如 MySQL 和 Oracle）或文件服务器中（文本、图片和视频文件），或是以电子健康记录的标准形式存在，如健康度 -7 (HL7) 信息。

**结构与知识相分离。**使用彼此独立的模型来表示数据结构和临床知识有几个好处。数据存储模型定义了数据存储实体，表示数据存储的语义，和所在域无关。域模型则表示临床知识和由数据存储模型所定义的通用数据结构的约束条件。通过明确域表示，域模型获得了对该域的系统独立规格。因此，当临床知识发生变化时无须更改软件，整个系统的鲁棒性变得更高。

**映射。**CHISTAR 数据整合与交换引擎会将指定源格式的数据映射至本地或全局域模型上。可能实现该映射的策略有三种：

- 单个全局域模型对所有数据源
- 单个本地域模型 每个数据源
- 多个本地域模型对每个数据源和共享的全局数据源

在第一种方式中，一个域模型对所有数据源表达共享的语义，这使域模型非常容易随着数据源的增加而变化。而且，要为全体数据找到一个最小化的域规格（或本体）可能会极其困难。

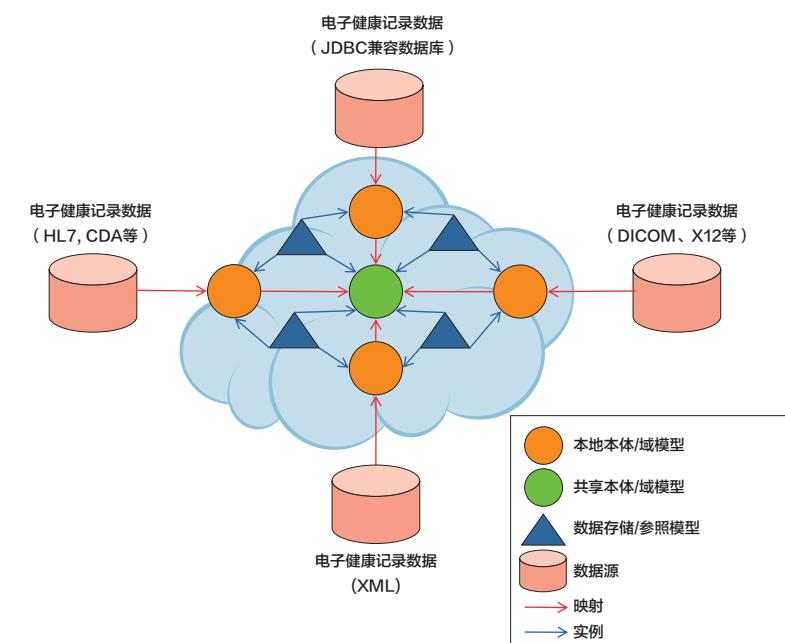


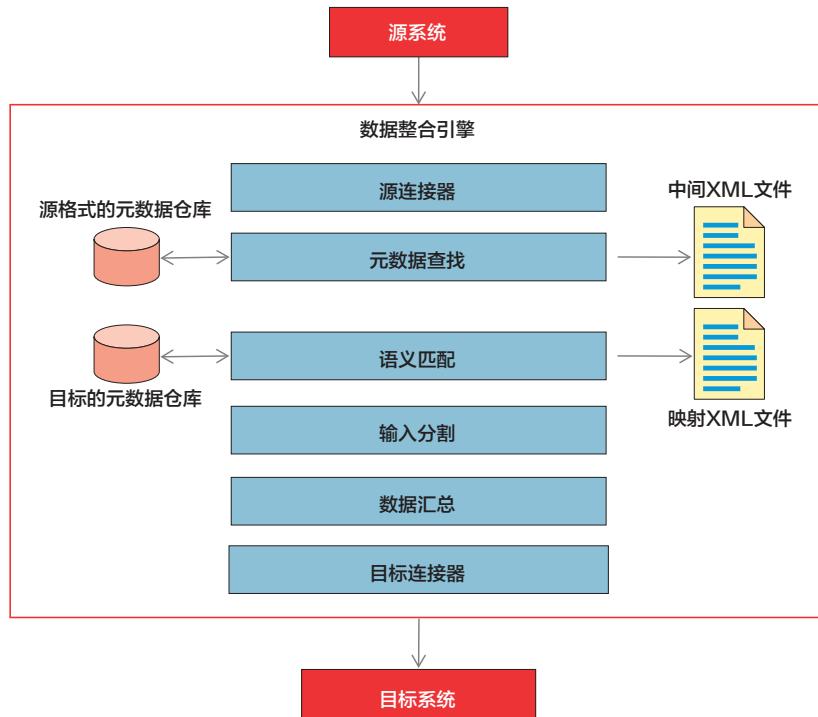
图 2. 推荐的数据整合方案。域模型是应用域的概念性展示，数据存储模型是数据存储的逻辑结构，映射在数据源和域模型之间建立关系。使用两套彼此独立的模型（数据存储模型和域模型）避免了因临床知识的变化而更换软件。CDA：临床档案架构；DICOM：医用数字化成像与通信；JDBC：Java 数据库连通性；HL7：健康度 7

在第二种方式中，每个数据源都有单独的（本地）域模型，这简化了域模型的结构。但是，通用本体的缺失让域模型间的比较变得更复杂，在它们之中定义映射变得更加困难。

第三种方式是前两种方式的混合。每个数据源都在一个（全局）域模型上拥有自己的域模型。全局域模型不易受

新增数据源的影响而发生改变，也更容易找到最小化的域规格（本体），因为与不同数据源相冲突的域知识会与每条数据源的本地域模型彼此分开。

**从源至目标的移动。**图 3 给出了数据整合引擎的架构。支持的标准包括 HL7、临床文档架构（CDA）、连续性护



**图 3.** 数据整合引擎。整合包括与外部系统的连接，解析输入文件来识别源数据格式，将其与储存在元数据仓库中的源数据格式进行语义匹配，对输入进行分割以利用并行机制，汇总数据并转换为目标系统的数据格式，最后连接至目标系统，将结果写入存储

理记录（CCR）、医用数字化成像与通信（DICOM）、X12、连续性护理文档（CCD）、XML、国家处方药物委员会（NCPDP）、电子数据交换（EDI）、定界文本以及ASCII源码。

一旦资源连接器就位，数据整合引擎就会执行元数据查找来发现源数据元素的语义。查找过程包括解析输入文件以及查找源数据的元数据仓库，检索全部源数据文件的数据元素的语义。数据整合引擎会为其支持的全部类型的数据进行元数据仓库的维护。

元数据查找过程由数据驱动，并生成一个 XML 中间文件，文件中包括所有源文件的数据元素，以及数据仓库查找的注释。XML 中间文件消除了源数据语句，并连同注释一起，保留了数据元素的层级和属性。

为了进行数据整合的下一步：语义

匹配，本框架将搜索目标格式的元数据仓库，并检索一份对中间文件中每一个数据元素的候选映射列表。为了引导搜索，本框架使用中间文件中源数据元素的注释。语义匹配可以是自动的，也可以是手动的。在自动匹配中，本框架会为全部源数据元素保留最相似的候选映射。

本框架将输入分割为并行的数据导入，并使用写入 MapReduce 并行编程模型（后文介绍）的作业（job）来汇总数据，并将其转换为目标格式。最后，目标连接器将数据文件写入 HDFS 存储中。

## 数据访问

数据访问层提供对云端医疗数据队列和检索的应用程序编程接口。数据访问层基于 Pig、Hive 和 hQuery 等开源技术建立。

Pig 是大型数据集的分析平台，由一

套表述数据分析程序的高阶语言和对这些程序进行评估的基础设施组成。应用开发者可以用 Pig Latin（Pig 使用的语言）写入程序脚本，Pig 的编译器会生成 MapReduce 程序序列，实现并行的数据处理。

Hive 在 Hadoop 上提供数据仓库基础设施，Hadoop 让数据查询和对存储在兼容 Hadoop 文件系统中的大型数据集进行分析变得更加便利。Hive 使用与 SQL 类似的 Hive 查询语言（HQL），能够查询 HDFS 或 HBase 中的数据。hQuery(<http://projecthquery.org>) 是一种对医疗数据进行分布式查询的开源框架。

HCatalog 为 Hadoop 提供元数据服务，建立在 Hive 元存储之上，并可以围绕元存储添加更多的层级。由于 HCatalog 可以为 Pig、Hive 和 MapReduce 提供共享的架构和数据模式，所以开发者无须担心数据在云端的位置。更重要的是，HCatalog 将数据分析应用与架构、位置或格式的变化分离开来，所以即使这些属性发生改变，开发者也不必重新编写数据分析应用。

## 数据分析

我们的三级架构通过数据分析层，在云架构上支持多种多样的数据分析算法。分析层使用 MapReduce 来构建医疗数据分析作业。数据分析层可以对云端收集到的大量医疗数据进行高效的数据分析。<sup>4</sup> 数据分析层基于 Hadoop 构建，Hadoop 框架提供 MapReduce 并行数据处理模型的开源实现。MapReduce 的处

理过程分两个阶段：映射（map）和规约（reduce）。在映射阶段，MapReduce 从 HDFS 等分布式文件系统中读取数据，在数据簇的运算节点中对读取的数据进行分区，然后将数据送至节点作为键值对。MapReduce 会独立处理每项输入记录，并将键值对作为中间数据存储在执行映射任务节点的本地磁盘上。

当全部映射任务完成后，开始进入归约阶段，包括对相同键值的中间数据进行汇总，也可以选择合并任务，在映射任务的输出传送给归约任务之前，就对相同键值的中间数据进行汇总。

## 支持高级医疗应用

从我们的三级架构功能中可以看出，数据整合和分析的功能强大，可以让开发者利用所需的工具对大量多样化数据进行开发，从而将医疗应用提升到更高的等级。电子健康记录系统包括个人级实验室检测结果以及诊断、治疗和人口统计学数据。尽管电子健康记录是设计目的，是实现病人和医疗服务商的临床交流，但记录中的数据也可用于种群级的健康监测、疾病的监测和爆发预测，及公共健康情况的掌握。

由于电子健康记录的数据不断更新，利用整合多个电子健康记录系统的框架可以让应用准确有效地预测疾病的爆发。我们的三级框架可以支持的更高级应用的实例包括：

- 流行病学监测，预测疾病的爆发

➤ 基于类似病人的决策智能，可以分析电子健康记录的数据，从中提取出与指定的目标病人最相似的一组病人记录。

➤ 不良药品事件预测，根据其他病人对药品的不良反应，预测哪些病人对某种药物产生不良反应的风险最高。

➤ 医疗纠纷调解，可以监测病人用药清单上的遗漏，识别出病人可能服用单上没有的药品。

➤ 预后分析，在类似病人预后数据的基础上，为病人预测所患病的可能的预后结果。

## 个人健康记录应用

图 4a 是一款个人健康记录应用病人概要信息页面的屏幕截图，把来自多个健康信息技术系统中的数据进行了统一展示。

## HealthMapper 应用

图 4b 是 HealthMapper 应用的屏幕截图，该应用根据地址和邮编将病例分组。为了将报告病例编组，我们每小时或每天将该应用离线运行一次，将汇总结果储存在 HBase 中。HealthMapper 会显示汇总结果。

该应用可以创建范围查询，显示特定时间段内的报告病例。HealthMapper 使用框架的数据访问层提供的编程接口和流行病学服务，对病人健康记录进行查询和分析。这一应用表明，使用我们推荐的三级框架，对整合自不同的健康 IT 系

## 使用实例评估

为了展示我们的技术栈（信息应用生成器、CHISTAR 中间件和三级框架）的效果，我们开发了一款同时整合了 100

使用单独模型分别表示数据结构和临床知识，让系统的鲁棒性更高，因为域知识的改变不会影响语义。

名病人数据的个人健康记录应用，还开发了 HealthMapper 应用，该应用可以通过对大量电子健康记录的分析，确认指定时间段内来自指定地区内的病例报告。为了部署我们的技术栈，我们使用了亚马逊弹性计算云（EC2）基础设施。

统的大量医疗数据进行分析，是可能实现的。

## 反应时间

图 5a 中记录了个人健康记录应用在三组应用部署配置下的平均反应时间，



图 4 使用推荐技术栈开发的应用成果。(a) 个人健康记录应用屏幕截图，图中可以显示来自多个不同结构的健康 IT 系统的数据。(b) HealthMapper 应用屏幕截图，显示了对大量数据的分析能力

以及同时供 100 人使用的一系列个人健康记录应用的反应时间。反应时间横向和纵向都有明显改善。

图 5b 中给出 HealthMapper 进行离线数据分类的时间。5H ( 大型 ) 组可以在短短几分钟内对 1000 万条记录进行分类。这表明，如有必要可以每小时执行一次分类作业。

**对** 来自不同医疗供应商（不仅是传统医疗机构和医生，还包括牙医、护士、理疗师和精神科医生等）的健康数据进行整合，不仅可以改善现有的就医状况，促进疾病预防的研究，还可以改善公共卫生问题。从短期来看，健康数据整合可以改善医疗的协调性，减少由数据缺失或不完整所造成的临床失误，减少重复检测，通过更加及时准确的协同用药提高对病人的安全性。从长期来看，健康数据整合将打开公共健康研究和人群监测的渠道，从而准确找出健康问题。

我们推荐的技术栈（包括信息学应用生成器、CHISTAR 中间件和三级框架）

共同为开发者提供了大规模数据整合和分析的工具，对应用开发具有显著的效益。

与客户端 - 服务器架构相比，本架构下的应用开发速度更快、成本更低，因为数据的整合、存储和分析均使用开源的云架构技术。

采用我们的三级框架构建的应用也能带来更多的合作机会，因为这些应用

的可达性非常高，用户可以在任何地点使用互联网连接安全登录系统。一致的数据表示、数据访问和对整合数据的解释也能够增进合作。

CHISTAR 中间件的安全功能，如角色访问控制可以让医生和专家在安全的环境下更自由地合作，因而提高了治疗的连续性。

另一个重要的效益是更强的可扩展

## 关于作者

**阿什迪普·巴哈 (Arshdeep Bahga)** 是佐治亚理工学院的研究员，他的研究领域包括云计算、大数据、数字信号和嵌入式系统。阿什迪普从佐治亚理工学院获得电气与计算机工程学硕士学位。联系方式：arshdeep@gatech.edu。

**维杰·K·马蒂塞蒂 (Vijay K. Madisetti)** 是佐治亚理工学院电气与计算机工程系教授。他的研究领域包括数字信号处理、嵌入式计算系统、芯片设计、无线电与电信通讯系统等。马蒂塞蒂在加州大学伯克利分校获得电气工程与计算机科学博士学位。他是 IEEE 会员，也是佐治亚理工学院驻印度机构的执行主任。联系方式：vkm@gatech.edu。

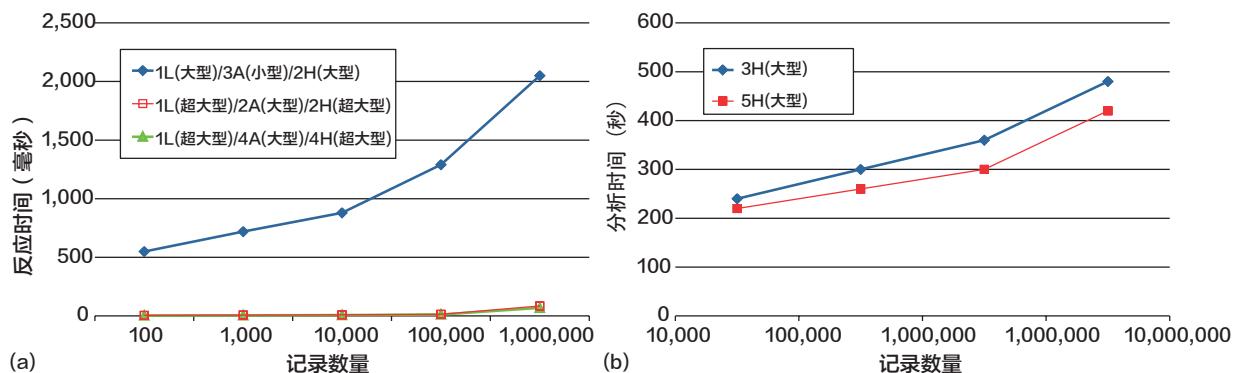


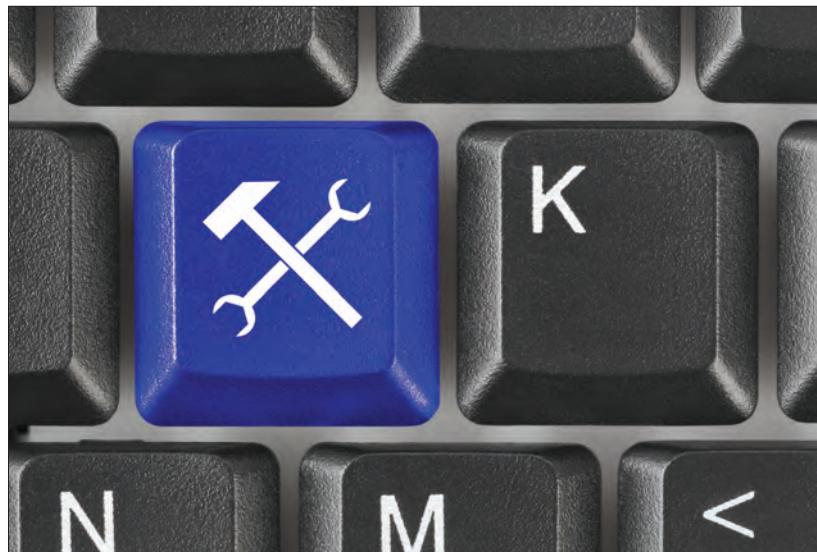
图 5. 应用反应时间。(a) 个人健康记录应用在最多 100 万条病人记录, 供 100 人同时使用下的平均反应时间。(b) HealthMapper 应用对最多 100 万条病人数据进行离线分类所用的时间。#L: 运行负载平衡器和网页服务器的实例数量; #A: 运行应用服务器的实例数量; #H: 运行 Hadoop/HBase

性。采用三级架构开发的云应用的可扩展性要高于客户端 - 服务器电子健康记录。随着整合的数据量的增加和新用户的加入, 这类应用使用的计算资源可以随需求的增加而增加。采用三级架构开发的应用能够实现横向扩展 (扩大规模) 和纵向扩展 (提升容量)。三级框架使用 HBase 存储, 可以通过添加节点来自动、线性地扩大存储空间。

最后, 采用三家框架开发的云应用降低了基础设施建设和运行成本。客户端 - 服务器应用需要一个 IT 专家团队负责安装、配置、测试、运行、安全和软硬件升级。而在云应用中, 所有这些功能都由云服务商来解决。C

## 参考文献

- A. Bahga and V. Madisetti, Cloud Computing: A Hands-on Approach, CreateSpace, 2013.
- A. Bahga and V. Madisetti, “A Cloud-Based Approach for Interoperable Electronic Health Records (EHRs),” IEEE J. Biomedical and Health Informatics, vol. 17, no. 5, 2013, pp. 894–906.
- C. Hajat, “An Introduction to Epidemiology,” Methods Molecular Biology, Jan. 2011, pp. 27–39.
- A. Bahga and V. Madisetti, Internet of Things: A Hands-on Approach, CreateSpace, 2014.

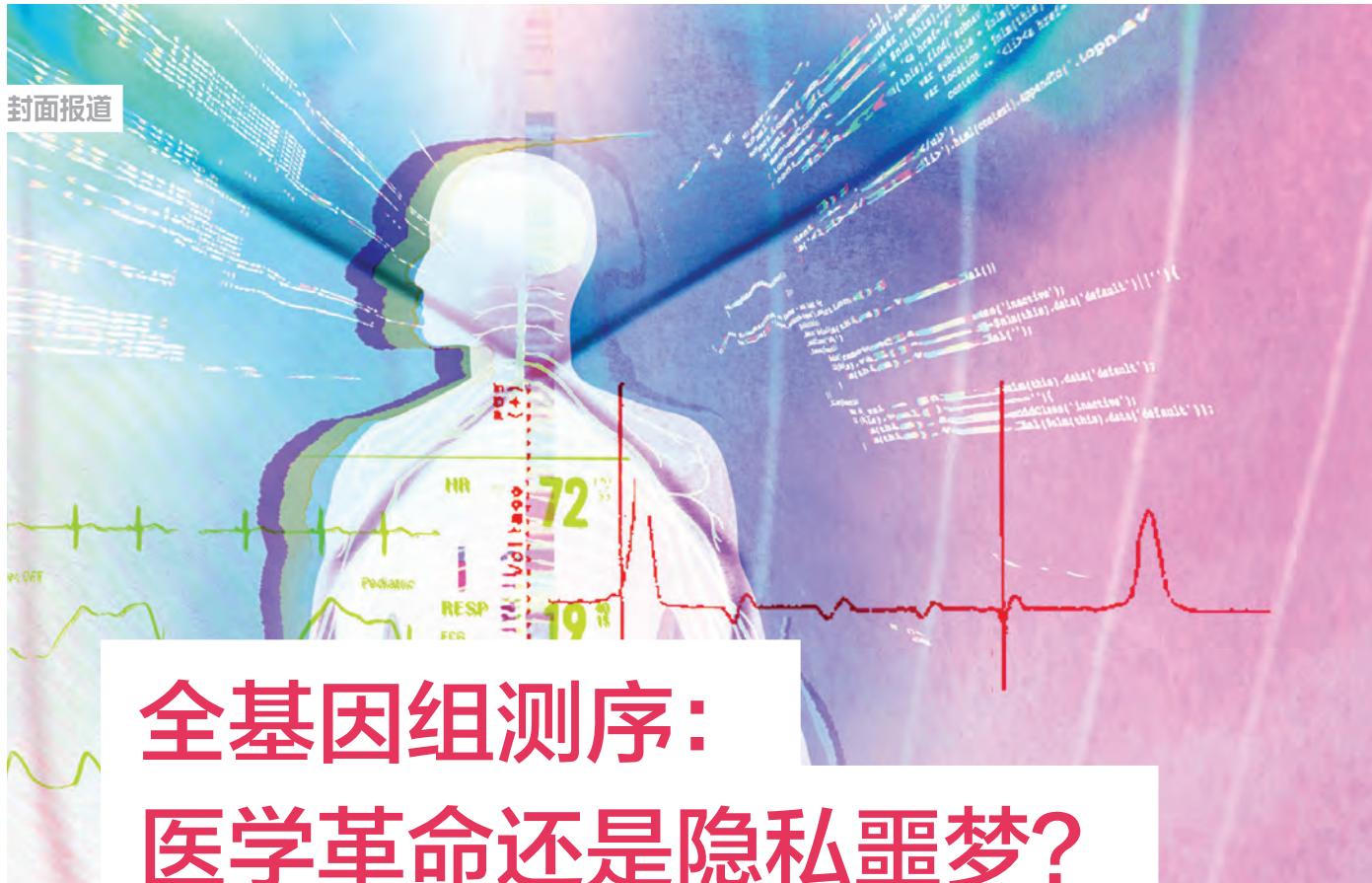


听听Diomidis Spinellis说什么  
播客：“交易工具”

[www.computer.org/toolsofthetrade](http://www.computer.org/toolsofthetrade)

Software

IEEE computer society



# 全基因组测序： 医学革命还是隐私噩梦？

厄玛·艾迪 (Erman Ayday)，毕尔肯大学 (Bilkent University)

埃米利亚诺·德克罗斯托法罗 (Emiliano De Cristofaro)，伦敦大学学院 (University College London)

让-皮埃尔·于博 (Jean-Pierre Hubaux)，瑞士洛桑联邦理工学院 (EPFL, Lausanne)

吉恩·丘迪克 (Gene Tsudik)，加州大学欧文分校 (University of California, Irvine)

很快，全基因组测序就能让许多人都负担得起了。

但棘手的隐私和伦理问题会妨碍它的推广，

并阻挠基因组学在医疗领域的大规模应用，延缓可能出现的医学进步。

**在**过去的十年间，全基因组测序 (WGS) 从未演化成可以用来获得个人整个基因组的现实技术。每个基因组序列都包含着大量的信息，这些信息能够显著地推动对疾病的认识、治疗和预防。因此，全基因组测序有潜力给医疗带来革命性的变化。

但是，基因组也包含着可以唯一地识别一个人的敏感信息。当技术进步最终使得一般大众能负担得起全基因组测序的时候，人们需要确保自己基因组信息访问权的安全。例如，谁来负责存储数字化的基因组，这些信息存在哪里？该如何控制访问权限，才能让任何人都无法无意或有意地把基因组信息泄露给第三方？怎样才能防止医疗机构的服务合作伙伴在医学研究和个性化医疗之外使用基因组信息？

随着单个基因组测序的价格下降到 1000 美元，这些问题变

得越发紧迫。新一代测序平台无论是通量的增加还是成本的降低都已经凌驾于摩尔定律之上了。因此，我们有充分的把握预测，几年内，在发达国家多数人都可以获得自己基因组的数字化信息。这些数据的用途很多，例如个性化医疗、亲子鉴定等等。Knome 和 Illumina 这样的商业实体已经在提供根据原始基因组数据生成报告的服务，医生可以利用这些报告指导治疗。

但是，如果对基因组和医疗之间的复杂互动缺乏更深刻的理解，全基因组测序的应用是有限的。要在这样的研究中取得进展，需要愿意分享自己基因数据的病人（或志愿者），而这样的许可则会引发对隐私保护、数据使用是否合乎道德以及法律权利的担心。例如，在个人基因组计划 ([www.personalgenomes.org](http://www.personalgenomes.org)) 中，参加者同意把自己的基因组数据和其他个人信息公开发布在互联网上。这样的实验性项目可以让我们一窥未来对处理大规模基因

## 保护基因数据的努力

**过**

去的几年中，基因隐私研究快速发展，现在分化成了 4 个主要类别：

- >> 字符串搜索和比较
- >> 聚合数据的发布
- >> 比对原始基因数据
- >> 基因组数据的临床应用，例如用于个性化医疗

第一个类别的研究的是医疗工具和私有字符串比较在保护隐私的亲子鉴定、个性化医疗和基因相容性测试中的应用。<sup>1</sup>最近，研究者扩展了此类研究的范围，把 GenoDroid 工具包<sup>2</sup>也包括进来了，这种工具可以通过智能手机进行亲子和血统鉴定。

在第二个类别中，研究者重点关注的是聚合基因组数据发布时的隐私风险。<sup>3</sup>其他人探索了差分隐私保护在聚合基因组试验统计数据发布中的应用。<sup>4,5</sup>他们的目标是，确保两个子还有个体数据差异的基因组数据库的统计特征不可区分。这样一来，一个基因组数据集发布的结果不会暴露数据集中存在某个特定个体。

第三类的研究者则在为 DNA 片段映射（拿数百万个短序列与一段参考 DNA 序列进行比对）寻找安全和高效的算法。这个方向上的最近一次尝试是在混合（公共和私有）云环境中进行的<sup>6</sup>。在这个研究中，作者把计算密集型的操作步骤外包给一个公共（不受信任或是商业的）云；他们建议把敏感和运算量较轻的计算托付给私有（可信任）云，以保护敏感 DNA 信息的隐私。

最后一类研究是在医学试验和个性化医疗中保护病人的隐私。一种方法是，用同态加密和安全多方计算在这些场合保护病人的基因组数据。<sup>7,8</sup>

一些努力的成果已经用在了实际的基因组测试中。不过，现在很难预测未来基因技术应用的范围和复杂性：可能有一些测试的运算过于复杂，无法在个人设备上进行，或者基因测试可能会牵涉到多个基因组。所以，我们预计，随着研究者不断取得新发现并且为满足新需求改变研究的重点，基因组数据保护的范围和性质也会相应发生变化。与此同时，正在进行的努力是通向基因组数据保护多方面挑战的解决方案的重要跳板。

### 参考文献

1. P. Baldi et al., "Countering GATTACA: Efficient and Secure

Testing of Fully-Sequenced Human Genomes," Proc. 18th ACM Conf. Computer and Communications Security (CCS 11), 2011, pp. 691–702.

2. E. De Cristofaro et al., "Genodroid: Are Privacy-Preserving Genomic Tests Ready for Prime Time?" Proc. ACM Workshop Privacy in the Electronic Society (WPES 12), 2012, pp. 97–108.

3. X. Zhou et al., "To Release or Not to Release: Evaluating Information Leaks in Aggregate Human-Genome Data," Proc. 16th European Conf. Research in Computer Security (ESORICS 11), 2011, pp. 607–627.

4. F. Yu et al., "Scalable Privacy-Preserving Data Sharing Methodology for Genome-Wide Association Studies," J. Biomedical Informatics, Feb. 2014, pp. 133–141.

5. A. Johnson and V. Shmatikov, "Privacy-Preserving Data Exploration in Genome-Wide Association Studies," Proc. 19th ACM Int'l Conf. Knowledge Discovery and Data Mining, 2013, pp. 1079–1087.

6. Y. Chen et al., "Large-Scale Privacy-Preserving Mapping of Human Genomic Sequences on Hybrid Clouds," Proc. 19th Network and Distributed System Security Symp. (NDSS 12), 2012; www.informatics.indiana.edu/xw7/papers/ndss2012.pdf.

7. E. Ayday et al., "Privacy-Preserving Computation of Disease Risk by Using Genomic, Clinical, and Environmental Data," Proc. Usenix Security Workshop Health Information Technologies (HealthTech 13), 2013; www.usenix.org/conference/healthtech13/workshop-program/presentation/ayday.

8. E. Ayday et al., "Protecting and Evaluating Genomic Privacy in Medical Tests and Personalized Medicine," Proc. ACM Workshop Privacy in the Electronic Society (WPES 13), 2013, pp. 95–106.

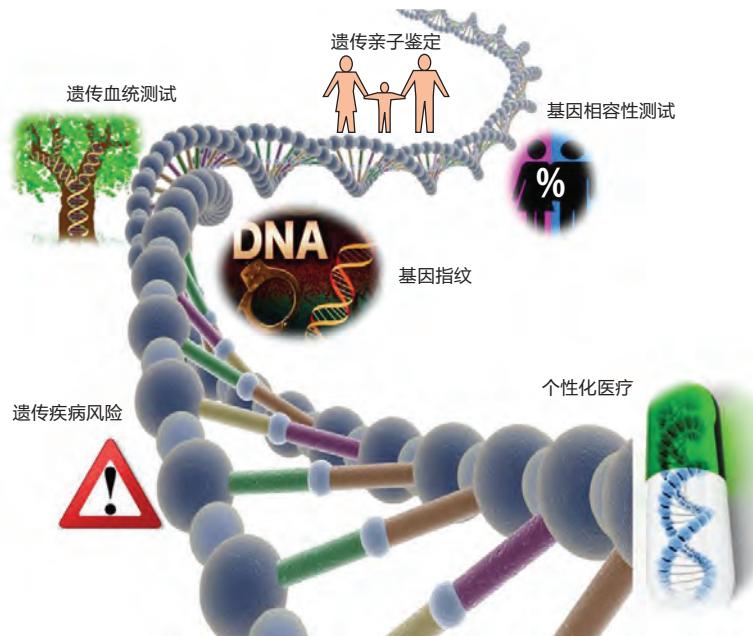


图1. 基因组学的应用。全基因测序使得个性化医疗成为了可能，还能帮助测试遗传病风险和血统

组数据的担忧。

DNA 测序进一步加剧了已经因社交媒体和个人健康记录 (PHR) 而广受关注的数据泄露和滥用问题。基因组代表一个人的生物身份，因此包含了关于这个人血统的丰富信息。把这些基因组数据和个人所处环境和生活方式的数据集合起来，第三方可以推测出该人的表现型，包括患上某种身体或精神疾病（例如阿尔茨海默氏病、癌症或精神分裂症）的可能性。

如果基因组信息泄露了，撤销或替换某个人的 DNA 序列都是不可能的，这会给依赖于准确基因组信息的应用领域带来严重影响。执法和医疗领域对 DNA 分析的使用已经引发了伦理问题，例如怎样保证基因组信息的完整性。

在研究者解决这些问题之前，个性化医疗那备受期待的美妙前景暂时难以成真。

## 基因组学 101

人类基因组由双链 DNA 分子编码，这些 DNA 分子包括两条互补的多聚物链。每条链都有一列核苷酸，用字母 A, C, G 和 T 代表。技术人员从人的唾液、头发、皮肤或血液等部位提取 DNA 样本，然后再提取出供测序用的遗传材料。测序出的基因组就是一条特异的字串，长度是 32 亿个字母对 (A, C, G 和 T 的组合)。

科学家用参考基因组来代表人类的基因组，占人类 DNA 序列的 99.5%。剩下的 0.5% 代表了人类个体间的基因变异。尽管看起来 0.5% 只占了参考基因组的一小部分，但它也对应了几百万个核苷酸。

遗传变异有几种形式，最常见的就是单核苷酸多态性 (single nucleotide polymorphism, 简称 SNP, 发音“snip”）。简单地说，一个 SNP 就是基因组序列上的一个位点，不同的人在这个位点上的核苷酸不同。例如，从两个人身上提取出的两段 DNA 片段分别是 AAGCCTA 和

AAGCTTA，第 5 个核苷酸分别是 C 和 T。

研究人员已经从人类基因组中确认了约 5000 万个独特的 SNP。<sup>1</sup> 随着越来越多的人同意进行测序，这个数字会变得更加确定。

SNP 可以帮助决定个体的患上某种疾病或失调的倾向性。例如，最近的全基因组关联分析 (genome-wide association study) 表明，3 个基因的 10 个 SNP 可以导致阿尔茨海默症的易感性。<sup>2,3</sup>

独立的 SNP 有时候会导致连锁不平衡 (linkage disequilibrium, 简称 LD)<sup>4</sup>——在两个或多个基因座上的等位基因中出现的非随机联系。这些基因座起源于单个祖先染色体，所以连锁不平衡有可能通过其他 SNP 暗示某个 SNP 的核苷酸。这一关系显然让隐私保护变得更复杂。

## 个性化医疗及其他

全基因组测序将带来预测 (predictive)，预防 (preventive)，参与 (participatory) 和个性化 (personalized, 和以上三者全称 P4) 医疗的新时代<sup>5</sup>，也让图 1 中的应用变成了可能。P4 医疗代表了保健领域的重大变革<sup>6</sup>，它不再像现有的试错治疗，因为它将用药方式和病人精准的遗传组成联系到了一起。P4 应用包括评估疾病和治疗的风险，亲子鉴定和始祖分析，以及评估与潜在伴侣间的遗传互补性，降低把遗传病传给后代的概率。

## 药物基因组学

实验表明，某些特定的遗传突变会改变药物代谢，所以基因组测试可以有助于预测病人对某种药物的反应。相关的实验和测试属于药物基因组学（pharmacogenomics）——这是一门研究遗传变异如何影响个体对药物反应的学科。药物基因组学的例子包括为患有白血病的孩子检测 *tpmp* 基因的 SNP 突变，以及针对和家族乳腺癌及卵巢癌相关的 *BRCA1/BRCA2* 基因进行治疗前检测。

一般认为，确定药物反应的基因组检测将很快得到更广泛的应用。专家预计现在正处于临床试验阶段的 900 种癌症药物中，有三分之一很快会附上 DNA 或其他分子检测推荐后进入市场。<sup>7</sup>

一些计划已经开始支持药物基因组学。例如，范德比尔特大学的加强治疗和保健的药物基因组学资源计划（Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment）<sup>8</sup> 评估病人的遗传学特征，帮助医生决定哪种药物最可能发挥效果，避免了传统药物评估过程中漫长的试错阶段。在一个案例中<sup>9</sup>，这项计划的研究人员利用一位心血管病患的遗传指纹，选择了特定的降胆固醇药物，并成功地治好了他。相比于传统方法，治疗只用了很短的时间。

## 检测遗传疾病风险

廉价全基因组测序让个人可以直接获得自己的遗传信息，他们可以把这些信

息共享给网站，检测遗传疾病风险。这样的网站包括 23andMe。23andMe 可以用相对较低的价格提供始祖和疾病风险测试服务，包括 96 万个 SNP，尽管它现在还没提供全基因组测序服务。自 2013 年 11 月开始，美国政府停止了 23andMe 和健康相关的检测，等待 FDA 调查。不过，英国仍能提供这些测试。

除了直接面向消费者的服务之外，国家和地区也正试图将基因组引入临床应用。比如英国的 10 万基因组项目（100,000 Genomes Project，网址 [www.genomicsengland.co.uk](http://www.genomicsengland.co.uk)），以及瑞士洛桑大学医院的 biobank（[www.chuv.ch/biobanque/bil\\_home/bil-patients-famille/bil-la\\_bil.htm](http://www.chuv.ch/biobanque/bil_home/bil-patients-famille/bil-la_bil.htm)）。

尽管研究人员正热情地研究遗传学和个性化医疗的关系，生物医学专家也表达了他们的疑虑：基因绘图到底能在多大程度上预测某种疾病出现的概率？<sup>10</sup> 他们认为，尽管科学家已经找到了一系列和某些疾病关联的遗传特征，<sup>2</sup> 他们并不知道环境因素是否（以及多大程度上）也会起作用。

## 亲子和始祖检测

在获得了病人完整的基因组测序后，医生和检测机构可以在几秒内完成复杂的遗传检测。相比昂贵的体外测试，这些专门的计算算法可以在法律接受的范围内实现更快更准确的测试。

已经有商业实体开始提供始祖检测和谱系检测。在这些检测中，软件可以比较个人的基因组信息和某个族裔公开

的基因组数据，确定被检测的个人和这个族裔之间的关系。线上服务还能提供遗传互补性检测，评估受检夫妇的孟德尔遗传风险<sup>11</sup>——孟德尔遗传风险是指把遗传病传给后代的概率。

## 基因组数据隐私的威胁

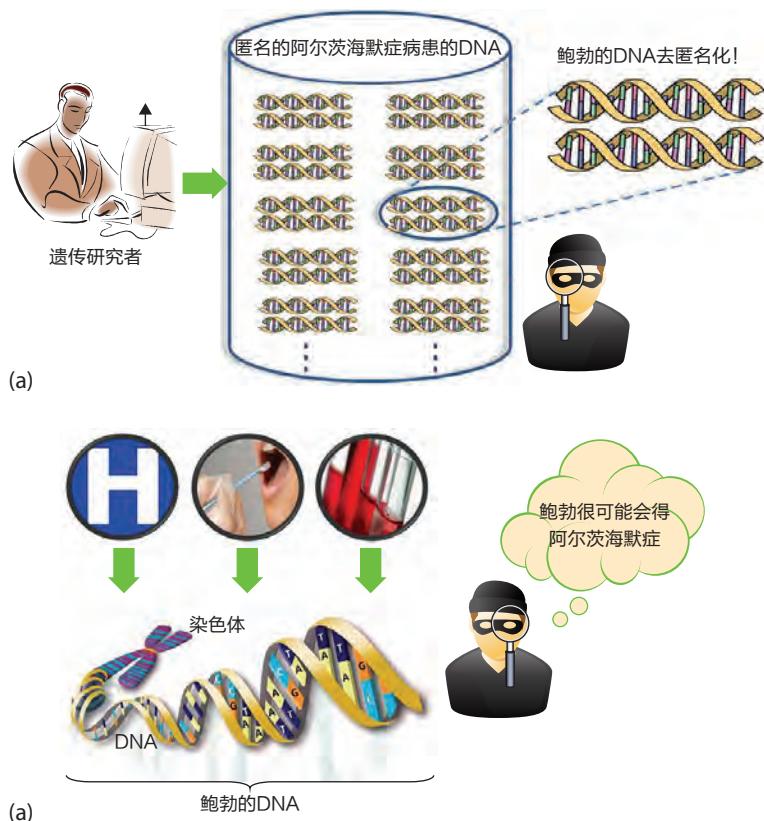
很多人对基因组隐私持怀疑态度，因为每个人都一直在留下自己的生物材料，比如头发、皮肤和唾液——这些都是证据，第三方甚至能在好几天后收集这些证据，完成 DNA 测序。但是，这些威胁主要针对目标个人或小群体，而不是针对大量的数字化基因组，比如研究数据库。

大量数字化基因组面对的威胁主要有两个，如图 2 所示。尽管已经有了一般性的法律可以用来保护数据隐私，我们仍然需要限制性更强的条款，应对特异性较强的隐私威胁。<sup>12</sup>

## 供体匿名化缺失

隐私保护最主要的传统方法是数据的去身份识别（data de-identification）和数据聚合（aggregation）。常见的去身份识别策略包括删除和屏蔽身份标识，比如姓名和社会安全号。不过，这些方法对基因组数据无效，因为基因组是一种终极身份标识。<sup>13</sup>

数据聚合是一种把人群数据结合在一起的方法，这种方法同样无效，因为已有足够的公开信息可以从某个案例研究中识别出个人，在某些情况下，还能复原部分基因组序列。例如，一项 2009



**图2** 人类基因组数据隐私面临两个主要的威胁。(a) DNA 的提供者在公开研究数据库中失去了匿名性(去匿名化), 以及(b)部分基因组数据泄漏让外部人员推测出敏感信息。图片来自美国能源基因组科学项目 (<https://public.ornl.gov/site/gallery/detail.cfm?id=398&topic=&citation=&general=dna&restsection=all>)

年的研究<sup>14</sup>表明, 即使是根据等位基因频率和已发表的论文计算出的测试的统计学(比如p值和r方), 也足以识别出参加遗传测试的人。一项2013年的研究<sup>15</sup>表明, 第三方可以从常见的谱系检测网站获得的信息, 再加上其他公开的个人数据, 就能对公开研究数据库中的DNA提供者重新进行身份识别(去身份识别的相反过程)。

### 数据泄露

因为两个亲缘相近的个人有着非常相似的基因组, 所以一个人的基因组被公开也有可能泄露给他亲缘关系较近的人的重要基因组信息。所以, 无论是自愿公开、意外公开还是恶意公开, 都存在问题。

有可能导致他人的身份泄露, 这让

基因组数据隐私变成了一个独特的问题。因为在大多数敏感的场合, 只有个人的数据处在风险中。而根据兄弟姐妹和子女的数量, 基因组数据泄露有可能影响一大群人。<sup>16</sup>如果没有考虑这种情况, 有可能导致严重的后果, 最近关于海莉耶塔·拉克斯(Henrietta Lacks)的基因组的争议就是证明。在将近50年前, 当科学家研究拉克斯的疾病时, 发现她的癌组织中细胞非常适用于生物遗传研究。科学家没有征得拉克斯家属的同意, 就提取了更多的细胞, 并开始利用这些海拉细胞(HeLa cell, 源自拉克斯的姓和名)进行研究。最后, 海拉细胞在遗传研究中变得非常常见, 所以拉克斯的家庭成员开始收到提取组织和血液样本的请求。在几次关于侵犯隐私的诉讼之后, 2013年,

美国国立健康研究院(NIH)同意给予拉克斯家庭一些海拉细胞使用的控制权。

让数据泄露问题的后果变得更加严重的, 是基因组的不变性和长期性。以个人可以修改密码、银行帐号, 甚至公钥证书。但他却不能修改自己的基因组。此外, 下一代人将遗传祖辈的大多数DNA, 所以遗传信息的泄露可能会变成无尽的诅咒。

### 隐私保护法律

很明显, 对隐私的担忧是阻碍大型人类基因数据库组建的重大因素, 可能导致基因组关联研究的延误或耽搁, 从而阻碍药物研发的进步, 进而阻碍医疗上的进步。对于越来越多使用DNA鉴定的执法部门来说, 保证基因数据的安全性和可靠性也是明显必要的。

现行法律只在某种程度上保护基因数据的私密性。1990年, 美国国家人类基因组研究所建立了伦理、法律和社会影响研究项目, 研究基因及个人、家庭和社群基因组研究对医疗技术进步的影响。2008年, 美国政府颁布《基因信息非歧视法》(GINA), 禁止医疗保险和雇佣领域基于基因信息的歧视。《医疗保险携带与责任法》(HIPAA)也为健康信息的保护和共享建立了框架。加州已经开始考虑针对DNA隐私立法。<sup>17</sup>同时, 欧洲的立法者们也在采取类似的预防措施。<sup>18</sup>

基因信息歧视并非新现象。早在1997年, 在著名的科幻电影《千钧一发》(Gattaca)中就涉及到了“基因主义”

的观点，即主张不同的人类性格和能力由基因决定的理论，并提出基因歧视的危害可能超过公然的种族歧视。

## 更严格的政策

尽管现行法律对基因数据的使用提供指导，但指导信息中却没有包括关于储存和处理数字化基因组信息的足够的技术信息。造成这种状况的一个原因就是，关于基因数据的安全和隐私问题，无论是个人基因组还是基因数据库中的数据，人们都还没能很好地理解。

隐私保护的支持者和消费者组织强烈支持实施更严格的立法来填补现有政策的缺口。最近，由美国总统组建的研究生物伦理学问题的委员会提交的报告对全基因组测序的发展进行了分析，<sup>19</sup>并强调了越来越高的隐私和安全风险，提出了几点关于隐私和安全问题的建议。

我们认为这些建议反映出对相关公开技术问题普遍的理解不足。比如，其中一条建议是非身份化(*deidentification*)，这明显不合适。这些建议同样没能解决几个重要问题。比如，防止暗中进行DNA测试的保护机制，任何基因数据的保护政策都必须承认知情同意权。相关政策中应该规定，权力机关或企业在对一个人的基因信息(如毛发或唾液样本)进行收集、分析、储存和共享之前必须获得这个人的书面许可，以保证没有人会成为未授权测序的受害者。

那些将隐私友好型政策视作基因研究障碍的人不会欢迎上述措施。科学家

一般通过对大量人群的DNA进行测序，来确定哪些基因与特殊的疾病有关。知情同意权的限制意味着他们不能重复使用大型基因组数据集来研究另一种疾病，而必须在每项研究之后将数据销毁，否则就要对之前所有的研究参与者进行回访，确保逐一取得新的授权来进行下一项研究。同时，由于有亲缘关系的人具有相似的基因组，所以研究参与者的亲戚可能也必须要作出授权才行。

## 关于基因数据保护与使用的指导意见

请求基因测序并很可能为之付费的个人应该拥有测序结果的所有权，就像对任何其他的个人健康信息一样。不过，基因组信息属于新型的个人健康信息，会产生大量单凭技术方法无法解决的问题。技术必须和法律与专业指导意见相结合，对基因信息的传播、储存、处理和最终处置进行管理。

### 储存与长期保护

对基因组信息的储存和保护存在几个重要问题：

- 基因组信息应该储存在个人设备上吗？为防止数据被侵入，需要具备哪些特殊的硬件安全功能？
- 基因组信息存储是否应该外包给云服务商？
- 基因组信息是否应该加密？如果

是，应该由哪些组织来生成并保管加密密钥？

虽然加密可能是解决上述许多问题的理想方案，但加密也有缺点。目前许多人看来强力的加密方案可能会逐渐变弱，但基因组信息的敏感程度不会降低。因此，现在无法对基因组信息进行解密的第三方可能在几年后就可以解密了。高级加密标准方案最高支持的密钥长度为256比特。虽然一些标准化组织和情报机构认为，这个长度在未来几十年内能确保安全，<sup>20</sup>但计算领域的突破或一些无法预料的弱点可能导致密钥被更早解密。

一种可选方案是假设信息不会被复制的情况下，定期对基因组信息进行重新加密。还有一种方法是利用秘密共享技术将其分割交给不同的服务商。不过，如何对信息进行有效重组，并确保服务商没有在基因组重建中互相串通成为问题所在。此外，服务商本身也必须要活的年头够长。

最后，加密不能防止已故多年的人的基因组数据遭到泄露，这会影响到其在世的后代。

### 可访问性

鉴于基因组数据高度敏感，一个人应该从不披露任何基因信息，这将阻止除个人安全设备之外的任何基因应用访问基因组信息。虽然这样看似理想，但这种限制只会在一些授信机构认证的一些标准化操作下才可能实现。比如，对遗传病的测试需要对基因组的某个大致位置与

某种知名模型进行比对，美国食品药品管理局可能会对这种模型及其参数进行认证，以保证受试者接受的是针对某种特定遗传病的合法测试，并会收到测试结果。受试者可以选择将结果保密。

其他关于数据访问的问题则更加复

## 目前许多人看来强力的加密方案可能会逐渐变弱，但基因组信息的敏感程度不会降低。

杂：

- 测序机构是否应该保有基因组数据的第三方副本？
- 受试者是否应该将基因组数据副本授权给医生或医保供应商？
- 是否有可能保证数字化信息的完整性和真实性？如果有，应该如何保证？
- 如果进行数据备份，应多久进行一次，保存在哪里？
- 是否有可能安全地删除基因组数据？
- 为享受更准确的技术，受试者是否应定期要求重新测序？

### 测试指导意见

为了有效取代体外测试，计算式基因测试必须要准确、高效，并可供遗传学专业人士使用。

#### 准确

计算式基因测试至少应该保证与体外基因测试相当的准确性。比如，计算式亲子鉴定应该提供与体外测试相同的信度，后者目前可被法庭采信。计算式

测试也应该争取担负起执行和信息输入正确性的保证。

#### 高效

计算式基因测试应该尽量将通信和计算成本降至最低。病人可能习惯花上数天来等待基因测试结果，但是在计算式环境中，在个人设备上运行时间过长可能会降低测试的实用性。

#### 可用性

计算基因组学检测很可能适用于一般人群，这会引发一些可用性的问题：

- » 用户需要知道基因组检测的哪些方面？
- » 应该让用户知道哪些关于检测和结构的信息？信息颗粒度有多大？
- » 个人对隐私的感知和顾虑是否和科学界的期望一致？

最后一个问题是特别复杂的。有些用户也许希望放弃他们的基因组隐私。比如，一个病人希望医生可以看到自己的基因组，让他们接受合适的检测，也许能将他们从致死疾病（比如癌症）中挽救回来。但是，同一个病人也许不希望在线服务上或制药公司看到自己的基因组信息。

这些意见大多只是有根据的推测，因为很少有研究专注于用户的问题。所以，研究的焦点应该放在探索用户上<sup>21</sup>，从而可以更深入地了解基因组隐私问题，并解这样一个开放式问题：如何有效地就潜在的基因组信息隐私风险进行沟通。

## 价

格适中的全基因组测序应用已经越来越广泛，这能带来很多激动人心的机会。但它也会引发隐私方面的担忧。从两方面研究全基因组测序需要遗传学家、保健服务提供商、伦理学家、法律制定者以及计算机科学家进行长期的合作。为此，我们协助组织了多学科的 Dagstuhl 研讨会。这一研讨会有助于讨论和基因组隐私相关的问题，在 2013 年举办<sup>22</sup>，并在 2015 年 10 月再次举办。我们还协助发起了国际基因组隐私研讨会，在 2014 年举办，并将和 2015 年 IEEE 安全与隐私大会（IEEE Symposium on Security and Privacy）联合举办。最后，我们还设立了 [www.genomeprivacy.org](http://www.genomeprivacy.org) 网站，为计算机科学家提供教程，让基因组隐私研究小组互相交流。

长期合作需要专门的资金支持。在

美国，基因组隐私处于机构资助范围的间隙里。例如，国立卫生研究院的基金完全覆盖了生物信息学和全基因组测序的伦理问题，但却只零星支持了几个基因组数据隐私方面的项目。国家自然基金会（National Science Foundation，简称NFS）的智能连接健康项目（Smart and Connected Health program）包括了需要计算机和健康科学合作的联合项目，但这一计划却未必能导致大范围的基因组隐私研究。

其他的美国基金到现在为止都没有明确基因组隐私问题。在欧洲，大量欧盟和欧洲国家资助的项目关注电子健康，有些会考虑数据保护，但却很大程度上忽视了基因组数据方面的隐私。此外，尽管大多数负责数据保护的官员都有很强的法律背景，但却缺乏计算机科学的专业知识。于是，他们就不让人感到意外地更多依赖立法而不是技术。

所以，我们的工作呼吁，需要研究合作来专门严肃地应对我们发现的隐私问题。解决了这些问题，就能让全基因组测序发挥所有潜力，在医学界掀起革命，个人和整个社会都能从中获益良多。■

## 参考文献

1. Nat'l Center for Biotechnology Information, "dbSNP," Dec. 2014; [www.ncbi.nlm.nih.gov/projects/SNP](http://www.ncbi.nlm.nih.gov/projects/SNP).
2. Eupedia, "Genetically Inherited Traits, Conditions, and Diseases," 2014; [www.eupedia.com/genetics/medical\\_dna\\_test.shtml](http://www.eupedia.com/genetics/medical_dna_test.shtml)
3. S. Seshadri et al., "Genome-Wide Analysis of Genetic Loci Associated with Alzheimer Disease," *J. Am. Medical Assoc.*, vol. 303, no. 18, 2010, pp. 1832–1840.
4. D.S. Falconer and T.F. Mackay, *Introduction to Quantitative Genetics*, 4th ed., Addison Wesley, 1996.
5. L. Hood and D. Galas, "P4 Medicine: Personalized, Predictive, Preventive, Participatory: A Change of View That Changes Everything," 2009; [www.cra.org/ccf/files/docs/init/P4\\_Medicine.pdf](http://www.cra.org/ccf/files/docs/init/P4_Medicine.pdf).
6. A. Weston and L. Hood, "Systems Biology, Proteomics, and the Future of Healthcare: Toward Predictive, Preventive, and Personalized Medicine," *J. Proteome Research*, vol. 3, no. 2, 2004, pp. 179–196.
7. A. Burke, "Foundation Medicine: Personalizing Cancer Drugs," 2012; [www.technologyreview.com/featuredstory/426987/foundation-medicine-personalizing-cancer-drugs/](http://www.technologyreview.com/featuredstory/426987/foundation-medicine-personalizing-cancer-drugs/).
8. My Drug Genome, "Using Genetics to Personalize Medication Treatment," 2014; [www.mydruggenome.org/overview.php](http://www.mydruggenome.org/overview.php).
9. K. Whitney, "PREDICT Helps Pinpoint Right Statin for Patient," *Vanderbilt Univ. Medical Center Report*, 4 Oct. 2012; <http://news.vanderbilt.edu/2012/10/predict-helps-pinpoint>.
10. G. Naik, "Gene Maps Are No Cure-All," *Wall Street J.*, 3 Apr. 2012; [www.wsj.com/articles/SB10001424052702304023504577319604245325644](http://www.wsj.com/articles/SB10001424052702304023504577319604245325644).
11. V. McKusick and S. Antonarakis, *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*, John Hopkins Univ. Press, 1994.
12. Y. Erlich and A. Narayanan, "Routes for Breaching and Protecting Genetic Privacy," *Nature Reviews Genetics*, vol. 15, no. 6, 2014, pp. 409–421.
13. N. Homer et al., "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays," *PLoS Genetics*, vol. 4, no. 8, 2008, pp. 1–9.
14. R. Wang et al., "Learning Your Identity and Disease from Research Papers: Information Leaks in Genome-Wide Association Study," *Proc. 15th ACM Conf. Computer and Communications Security (CCS 09)*, 2009, pp. 534–544.
15. M. Gymrek et al., "Identifying Personal Genomes by Surname Inference," *Science*, vol. 339, no. 6117, 2013, pp. 321–324.
16. M. Humbert et al., "Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic

## 关于作者

**厄玛·艾迪 (Erman Ayday)** 是土耳其安卡拉的毕尔肯大学的计算机科学助理教授。他在进行本文报道的研究时，还是瑞士洛桑联邦理工学院的博士后。他的研究兴趣是隐私、基因组学、信任和信誉系统和网络安全。艾迪在佐治亚理工学院获得了博士学位，他是 IEEE 和 ACM 的会员，他的联系地址是：erman@cs.bilknet.edu.tr。

**埃米利亚诺·德克罗斯托法罗 (Emiliano De Cristofaro)** 是伦敦大学学院的高级讲师（副教授）。他在进行本文报道的研究时，是施乐公司的帕洛阿尔托研究中心 (PARC) 的研究科学家。他的主要研究兴趣是隐私增强技术和应用密码学。德克罗斯托法罗在加利福尼亚大学欧文分校获得了网络系统博士学位，他的联系地址是：me@emilianodec.com。

**让-皮埃尔·于博 (Jean-Pierre Hubaux)**，是瑞士洛桑联邦理工学院计算机与通信学院教授。他的研究兴趣是隐私保护，特别是在移动网络和基因组学中。于博是在米兰理工大学获得工程学博士学位的。他是 IEEE 和 ACM 的会士，他的联系地址：jean-pierre.hubaux@epfl.ch。

**吉恩·丘迪克 (Gene Tsudik)** 是加利福尼亚大学欧文分校的计算机科学校长讲习教授。他的研究兴趣包括安全，隐私和应用密码学。丘迪克在南加州大学获得了计算机科学博士学位，他是 IEEE 和 ACM 会士，他的联系地址：gts@ics.uci.edu。

Privacy,” Proc. 20th ACM Conf. Computer and Communications Security (CCS 13), 2013, pp. 1141–1152.

17. H. Shen, “California Considers DNA Privacy Law—Academic Researchers Fear Measures Would Prohibit Work with Genetic Databases,” Nature, 18 May 2012; www.nature.com/news/california-considers-dna-privacy-law-1.10677.

18. Council of Europe, “Additional

Protocol to the Convention on Human Rights and Biomedicine, Concerning Genetic Testing for Health Purposes,” 2008; http://conventions.coe.int/Treaty/EN/Treaties/html/203.htm.

19. Presidential Commission for the Study of Bioethical Issues, “Privacy and Progress in Whole Genome Sequencing,” 2012; www.bioethics.gov/cms/sites/default/files/Privacy\_Progress508.pdf.

20. Nat'l Inst. Standards and Tech.,

“Cryptographic Key Length Recommendation,” 2014; www.keylength.com/en/4.

21. E. De Cristofaro, “An Exploratory Ethnographic Study of Issues and Concerns with Whole Genome Sequencing,” Proc. 8th Network and Distributed System Security Symp. (NDSS) Workshop Usable Security (USEC 2014), 2014; http://arxiv.org/abs/1306.4962.

22. K. Hamacher, J.-P. Hubaux, and G. Tsudik, “Dagstuhl Seminar on Genomic Privacy,” Oct. 2013; www.dagstuhl.de/en/program/calendar/semhp?semnr=13412.

征稿启示  
让自己出现在人工智能领域的最前沿吧！

在IEEE Intelligent Systems上发表你的文章

IEEE Intelligent Systems  
寻找所有人工智能领域的文章，它专注于将最新的研究付诸应用的开发工作。

投稿指南参见  
www.computer.org/mc/intelligent/author.htm.

最好的人工智能杂志 Intelligent Systems  
www.computer.org/intelligent



## Instant Access to IEEE Publications

Enhance your IEEE print subscription with online access to the IEEE Xplore® digital library.

- Download papers the day they are published
- Discover related content in IEEE Xplore
- Significant savings over print with an online institutional subscription

*"IEEE is the umbrella that allows us all to stay current with technology trends."*

Dr. Mathukumalli Vidyasagar  
Head, Bioengineering Dept.  
University of Texas, Dallas



Start today to maximize your research potential.

Contact: [onlinesupport@ieee.org](mailto:onlinesupport@ieee.org)  
[www.ieee.org/digitalsubscriptions](http://www.ieee.org/digitalsubscriptions)

 **IEEE**  
Advancing Technology  
for Humanity

# 环球科学 SCIENTIFIC AMERICAN

阿尔·戈尔 (Al Gore)

2007 年获诺贝尔和平奖  
1990 年和 1991 年在《科学美国人》  
上发表两篇文章，大声疾呼人类应该  
尽早行动，抵御全球变暖



比尔·盖茨 (Bill Gates)

美国微软公司创始人  
2007 年在《科学美国人》  
上发表文章，预言机器人产  
业的兴起



斯蒂芬·霍金

(Stephen Hawking)  
著名物理学家  
2010 年在《科学美国人》  
上发表文章，阐述他对物理  
学和我们这个世界的思考



1845 年创刊

152 位诺贝尔奖得主撰稿  
传承百年的科技媒体品牌  
科技精英分享智慧与见解  
的全球化平台

罗伯特·布鲁斯·梅里菲尔德  
(Robert Bruce Merrifield)

1984 年获诺贝尔化学奖  
1968 年在《科学美国人》上发表文章



安德烈·K·海姆

(Andre K. Geim)

2010 年获诺贝尔物理学奖  
2009 年在《科学美国人》上发表文章

**穆罕默德·尤努斯**  
(Muhammad Yunus)

2006年获诺贝尔和平奖  
1999年在《科学美国人》上发表文章，  
介绍用小额信贷帮助贫困人口



**朱棣文 ( Steven Chu )**  
1997年获诺贝尔物理学奖  
1992年在《科学美国人》  
上发表文章



**弗朗西斯·克里克**  
(Francis Crick)  
1962年获诺贝尔生理学或医学奖  
多次在《科学美国人》  
上发表文章

汇集 nature 和《科学美国人》  
独家版权内容  
全球数千科技精英撰稿  
为你报道最新科研趋势与最具  
前景的科技成果



《环球科学》是全球著名科技媒体《科学美国人》杂志中文版，创刊170年来，它始终站在科技最前沿，在美、德、法、意、日等15个国家出版同步发行，内容涵盖生物、医学、IT、电子、能源、经济等广泛领域，是科技企业、知识阶层与政府决策者首选的科技指南。

《环球科学》同时获世界权威科学期刊《自然》(nature)授权，独家刊载《自然》精华内容，第一时间呈现世界科技发展全貌。

您可采取以下方式订阅：

- ① 邮发代号：80-498
- ② 京东、亚马逊、当当网、天猫、淘宝、1号店等均有售
- ③ 杂志社官方淘宝店购买：  
<http://huanqiukexue.taobao.com>

**订阅热线**  
**010-57458982**

更多信息，请登录官方网站：  
[www.huanqiukexue.com](http://www.huanqiukexue.com)



# 分析学、机器学习 和物联网

赛斯·厄利 (Seth Earley), Earley & Associates

我们的世界正越来越紧密地连接在一起，再加上廉价传感器和分布式智能，将导致工业界产生重大的变革，制造出人类难以处理的大量数据。公司是否能迅速适应和变化，从而在商业竞争中维持自己的地位？人类如何理解和利用这些新的信息来源以及周围环境中的智能信息？

## 开发演变的技术

组织机构将需要合理搭建自己的内部数据仓库，从而利用新的数据源和数据流。在某些情况下，智能连接设备也会在某些循环中移除人工过程，这些设备将在需要的情况下自动作出决策，自我调整，校正方向，以及自我修复。在其他情况下，设备集合会成为可被新方法优化的系统，而不同系统组成的系统将会分享数据，成为数据和设备的生态系统。机器学习这个词描述了无数种挖掘数据含义的方法。当组织机构在为物联网（简称 IoT）作准备时，机器学习是解决方案中必备的部分，但传统的商业和数据分析技术同样

不可或缺。

物联网被某些人称为万物之网（Internet of Everything），这一领域正加速增长。高德纳咨询公司（Gartner）预计，到 2020 年，物联网中将有 260 亿台设备，物联网产品和服务的价值将达到 3000 亿美元。<sup>1</sup>GE 长期以来一直在工业互联网领域耕耘，它估计在接下来的 20 年，工业互联网将会让全球国内生产总值提升 10 到 15 万亿美元（没错，是万亿）。<sup>2</sup>工业互联网是指监控和优化工业设备（包括喷气机引擎、机车、动力涡轮机和生产流程）的原理和应用。

当然，目前在市场上有大量关于新出现技术的炒作。事实上，高德纳著名的“炒作循环”报告将物联网列在“期望夸大的顶峰”阶段（数据科学已经进入了“理想破灭的低潮”阶段）。<sup>3</sup>但是，虽然企业家热情如火，记者们对未来的前景也很期待，但是组织机构必须克服很多挑战才有可能利用相关的技术进步。

## 存在的挑战

组织机构必须关注：

- 在产品技术和 IT 方面理解企业能力的相对成熟度。
- 理解物联网可被合并的功能类型，以及新的性能将会在什么地方影响客户价值。
- 理解机器学习和预测分析模型的作用。
- 根据市场变化速度和竞争对手的相对敏捷程度重新思考商业模式和价值链。

下面我们会分析所有这些挑战的细节问题。

### 理解产品和 IT 的成熟度

可以从两个维度来考虑这个问题。产品组合有多成熟？是变化缓慢、逐步进化的传统产品类型，还是快速前进、更加复杂的技术生态系统？开矿设备从技术上来说很复杂，但是相比科学仪器来说，它的设备生命周期更长，进化速

度也更慢。然而，这并不意味着仪器公司能更好地为系统优化提供物联网设备。另一个需要考虑的因素是 IT 过程的成熟度。所有的组织类型都能从启动物联网中获得好处，但是这种变化的模型会有所不同。

现在从 IT 成熟度水平这一维度进行分析。举个例子，科学仪器的供应商也许在技术上更先进，但并没有很好的 IT 架构、IT 过程和控制力。而开矿设备的生厂商也许有非常成熟的内部 IT 过程。这也许意味着，科学仪器公司在升级现场仪器操作的功能时，有可能会考虑到物联网，但不会试图优化包括多种不同设备种类的实验室信息生态系统。（当然，作为成本中心的 IT 成熟度缺失也许不会转化为利润中心的 IT 成熟度缺失，但是很多组织机构在开发或拓展 IT 服务产品时，以已有的基本 IT 性能为基础。）挖矿设备的案例最近被一篇发表于《哈佛商业评论》上的文章讨论：Joy Global 是一家矿业设备的生产商。它可以利用自己在矿业操作相关的多个系统和流程方面的专业知识，监控、维护和优化自己生产的设备。<sup>4</sup>

#### 理解物联网的性能

下一个被考虑的问题是智能连接的产品中有哪些性能可以被利用？根据上面提到的那篇《哈佛商业评论》的文章，物联网的性能有 4 种类型：<sup>4</sup>

- 监控——传感器可以提供操作环境和产品使用及产品性能的数据；
- 控制——产品功能可被控制和个性化；
- 优化——来自监控和控制的反馈环路可以改进效率，改善性能，还

能增进预防性维护、诊断和修理；  
• 独立化——监控、控制和优化可以允许独立操作、与其他系统协调、与环境交互、个性化、补给、自我诊断和修复。

这些水平上的性能有可能重新定义供应链，并重构价值链。我们不应该再把产品的功能看成固定的东西，我们要看到产品的灵活性和适应性。当产品智能化并被连到互联网上以后，它们变得更加多变，也可以根据用户需求的变化而变化。软件制造商认识到这点已经有好几年了。现在，物理物体已经变成了有软件驱动

让供应链中不再有分销成本和库存。

控制是更复杂的应用，层级在监控至上。我们可以监控设备操作，然后再通过控制设备的多个部件或多个系统，扩大人类干预范围的边界。以操作系统和机器（大部分功能是自动的）的人扮演的角色为例。人类引导操作，寻找边界条件、异常情况，以及在系统设计中预料之外的例外情况（或是成本效益）。随后，他们根据自己的判断作出改变、校正或调整。人类不需要和机器在一起，可能也不需要实时地进行监控（根据流程决定）。监控只是收集并处理数据（这些是对数据必要的处理）。控制是实时（或

## 我们不应该再把产品的功能看成固定的东西，我们要看到产品的灵活性和适应性。

功能的汽车和集装箱。这些水平的性能需要更复杂的数据分析方法——比如收集和应用数据，以及让算法自己应用数据并在同时学习。

所以，这些能力的第一层——监控，变成了实时机制，目的是更好地理解现场性能和用户需求，并提供新性能。这也意味着组织的传统产品与服务的边界被拓宽，也变得模糊。考虑一下现场设备，传统上这些设备由现场服务承包商（而不是生产商）维护。有了智能和监控后，设备可以在出现故障之前就通知制造商。日常维护可以成为制造商服务的一部分，而复杂的维修仍然可以由专门的制造商负责，如果组织可以接受相应的利润和物流的话。这种去中间化的过程可以应用到分销链上。设备可以自动提出补给需求，

接近实时）地将数据应用于设备操作中。组织需要作的战略决策：是否以及何时在提供的产品中增加控制性能，以及是否把这种控制能力作为服务或是允许客户可以对其进行操作。

性能的第三个层级——优化，可以拓展到单个物体、一队物体或是一个物体的生态系统的性能上，覆盖了多个制造商和技术。关于是否把提供的服务延伸到这些领域的战略决定依赖于价值链的知识水平和复杂程度，以及过程的边界。挖矿的例子说明，矿业制造公司 Joy Global 相比于流程生态系统视野较小的供货商，会有优势。例如，一家卡车制造商很可能在优化复杂的挖矿设备方面很不在行，但却能从优化自己的卡车产品（很可能还有其他制造商的卡车产品）中获益，如

果这个行业的动力学符合商业逻辑的话。

想要将优化过程拓展到独立的操作上，需要增强性能，允许与环境和其他系统进行限制更小的交互。独立化要求算法有更强的智能，从而可以处理一些计划之外的状况——也就是程序员和系统工程师没有明确设计的状况。独立操作需要使用可改编的机器学习方法，从而能让用于监控、控制和优化的核心算法能应对新情况。

## 理解分析学和机器学习

2014年11月，施乐帕罗奥多研发中心的迈克·库尼雅夫斯基（Mike Kuniavsky）曾在IDTechEx发表题为《物联网预测性分析的用户体验》（The User Experience of Predictive Analytics in the Internet of Things）的演讲。他认为几乎所有的功能已存于（或将很快存于）云端。数据和功能可以从任何地方通过多种设备获得。

运动手环可以通过iPhone或笔记本电脑可以在运动这种特定的情况下评估用户的生理健康数据。运动手环相当于物联网的传感器，同时也能提供获得数据和消费数据的方法。它还能通过软件功能覆盖其他设备（比如计步器）。这些设备提供的数据还可以帮助我们了解用户的习惯和爱好，而这又可以在升级功能和开发新特性时提供参考。如果聚集了大量用户，并结合其他数据库，就能产生新的理解，弄清楚流行病学数据，不同人群的活动水平，生活方式，以及人口统计学数据。这些信息对营销商、保健服务提供商、保险公司和政府机构来说都很有价值。（当然，我们需要负责考虑隐私和数据使用权限。）

机器学习算法可被用来基于数据模

式作出预测。例如，梅约医院（Mayo Clinic）的研究表明，活动数据和心脏病人的康复速度有关系。<sup>5</sup>

同样的机器学习和预测算法是很多连接的智能消费设备的基础。Nestde1恒温器就是这么个例子，它能在一天里的某个时间，利用数据模式预测某个特定的房间所需的气温。（另一个控制和优化的例子是在社区层面上的，发电站可以在住户同意的情况下，把数百到数千个Nest的设备调整几度，用电高峰期错开能量负载。）其他消费设备可以从声音模式中学习（例如Amazon的个人助理型设备Echo<sup>6</sup>），或是可以从更复杂的行为或活动模式中学习（例如捷豹路虎的监控系统，可以“依靠能让汽车学习、预测、检查和提醒车主的复杂软件，帮助驾驶员自动把任务委托给汽车，把精力更好地集中在驾驶上。”）

优化算法使用机器学习原理，从而利用与动态环境交互的传感器和智能设备的数据。在没有某些特定参数的情况下，这种多变的环境不能被精确预测。这些算法需要感知、响应和适应。例如，当汽车承担了更多驾驶员的责任时，他们将与更多的环境数据源交互（比如传感器、灯光，和其他汽车等等）。工业自动化、物流运输、电网和能源系统、交通管理、安全系统，以及其他“系统的系统”中的应用，将让机器直接和其他机器交流。此外，这些应用将帮助机器根据可以进化和适应的算法，解释数据流。这样，机器就可以在给定某些操作参数的情况下，达到理想的最终状态。

## 重新考虑商业模型和价值链

智能并联网的设备需要组织重新考虑自己应何时及如何在市场中创造价值，

以及创造出的价值会如何随着信息生态系统和竞争环境的变化加强或减弱。分析学可以帮助验证一些决策（例如，获得关于功能变化和新增服务及功能的实时使用数据）。然而，商业模式在新加入的公司和价值链结构的作用下，可能会发生剧烈的转型，让基于公司传统商业模型的分析方法失效。产品或服务可能基于因旧产品而效果不好的数据流，而不是基于从产品本身获得的收入。新的商业模式的含义也许可以远远超过产品本身，影响上游供应商或下游的消费者。



核心上来说，所有这些可能性都需要组织机构建立基本的能力，清洗内部数据，构建分析基础架构：数据保管（data curation），所有权和质量标准，统一的企业架构，有条理的整合系统，自动数据载入流程，以及成熟的数据分析专家。如果没基础没有搭建好和管理好，就很难迅速地对新的分析及数据管理功能作出反应。

因为物联网将基于数据流，也将基于复杂的方法，可以从信息中获得深入了解并将之整合进企业知识中，应用于价值创造过程。组织机构如果没有这样的能力，将在市场中落后，或是沦为低价值，低利润物品的提供者。数据被称为新的石油——如果扩大一下这个比喻，数据可以通过有分析能力的知识精炼厂炼制成高附加值的产品。组织机构现在需要在这样的基础设施上进行投入，这样它们就能在接下来的几年里准备好面对转化的、被颠覆的供应链和价值创造过程。信息的敏捷型将会是一个必须的核心能力。■

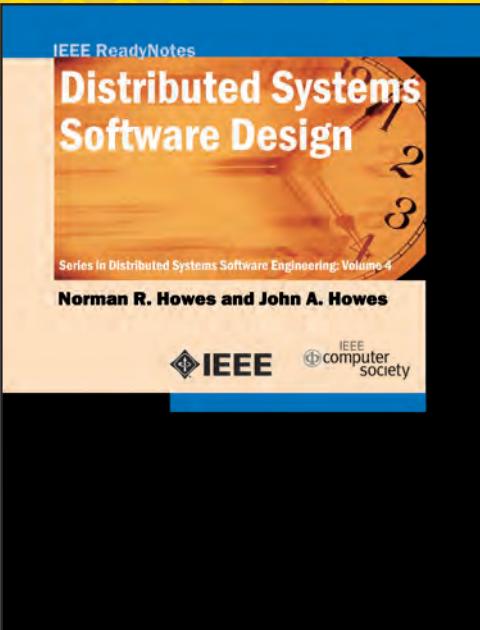
（本文内容来自IT Professional）



## 参考文献

1. "Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020," Gartner, 12 Dec. 2013; [www.gartner.com/newsroom/id/2636073](http://www.gartner.com/newsroom/id/2636073).
2. "Analyze This: The Industrial Internet by the Numbers & Outcomes," GE, 7 Oct. 2013; [www.gereports.com/post/74545267912/analyze-this-the-industrial-internet-by-the-retrieved-2014-11-30](http://www.gereports.com/post/74545267912/analyze-this-the-industrial-internet-by-the-retrieved-2014-11-30).
3. "Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business," Gartner, 11 Aug. 2014; [www.gartner.com/newsroom/id/2819918](http://www.gartner.com/newsroom/id/2819918).
4. M.E. Porter and J.E. Heppelmann, "How Smart Connected Devices are Transforming Competition," Harvard Business Rev., Nov. 2014, pp. 70-86.
5. D.J. Cook et al., "Functional Recovery in the Elderly After Major Surgery: Assessment of Mobility Recovery Using Wireless Technology," Annals of Thoracic Surgery, vol. 96, no. 3, 2013, pp. 1057-1061; [www.annalsthoracicsurgery.org/article/S0003-4975\(13\)01253-8/fulltext](http://www.annalsthoracicsurgery.org/article/S0003-4975(13)01253-8/fulltext).
6. D. Etherington, "Amazon Echo Is A \$199 Connected Speaker Packing an Always-On Siri-Style Assistant," Tech Crunch, 6 Nov. 2014; <http://techcrunch.com/2014/11/06/amazon-echo>.
7. M. Mendoza, "Jaguar Land Rover Develops Self-Learning, Intelligent Car," Tech Times, 17 July 2014; [www.techtimes.com/articles/10308/20140717/jaguar-land-rover-develops-self-learning-intelligent-car.htm](http://www.techtimes.com/articles/10308/20140717/jaguar-land-rover-develops-self-learning-intelligent-car.htm).

赛斯·厄利 (Seth Earley) 是 Earley & Associates ([www.earley.com](http://www.earley.com)) 的 CEO。他是知识过程和用户体验管理战略的专家。他的研究兴趣包括用户体验设计, 知识管理, 内容管理系统和战略, 以及分类学研究。联系方式: [seth@earley.com](mailto:seth@earley.com)。



**IEEE ReadyNotes**  
**Distributed Systems Software Design**  
Series in Distributed Systems Software Engineering: Volume 4  
Norman R. Howes and John A. Howes  
IEEE computer society

**NEW from IEEE CS Press**

**DISTRIBUTED SYSTEMS SOFTWARE DESIGN**  
by Norman R. Howes and John A. Howes  
Volume 4 in the Series in Distributed Systems Software Engineering. Discusses how to design certain types of distributed systems and how to specify these designs both informally and formally. Essential reading for all designers of safety-critical systems.  
Product ID RN0000023 • .PDF exclusive • 117 pp.  
Order .PDF (\$19):  
<http://bit.ly/12xaghP>  
Order other .PDF volumes in this series:  
<http://webstore.computer.org>

Editors: rick Kuhn, Us National institute of standards and technology, kuhn@nist.gov tim Weil, Us department of interior, trweil@ieee.org



## 索尼的遭遇也会降 临到我们头上吗？

沃尔特·豪泽 (Walter Houser) , 开放式 Web 应用程序安全项目 (Open Web Application Security Project)

2014 年，索尼、家得宝、JP 摩根、eBay 和塔吉特都曾因同样的原因无奈地出现在新闻报道里。此外，虽不像前几个公司的事件那么广为人知，Gmail、Mozilla、社区健康服务公司、纽约出租车、LexisNexis、美国在线和韩国征信公司等机构也遭遇了严重的数据泄露。看起来几乎每天都有新的企业成为受害者，攻击者带着海量的个人和公司信息逃之夭夭。这类新闻报道出现的如此频繁，以至于我们都已经对数据泄露感到疲劳了。

数据泄露如此普遍，给人的感觉是我们的数字基础设施已经濒于崩溃。越来越多的消费者在信息高速公路上，而不是现实世界中遭到抢劫。这样的抢劫在财务上造成的后果可能更加可怕，因为网上的受害者可能丢掉全部积蓄并失去良好的信用记录，小巷子里拦路抢劫的小混混可是做不到这些的。

虽然人们的反应往往是自甘倒霉而不是满腔愤怒，但有少数受害者会一直生活在愤怒和恐惧中：那些被网络犯罪

分子夺走了自己的身份控制权的人。这些可怜的人注定要用上几个星期的时间跟征信机构、讨债公司和执法人员周旋，试图为自己“受伤严重”的财务和声誉止血。

在数以千计内容丰富、充满诱惑的钓鱼电子邮件的轰炸下，有太多的人点击了里面的链接，让自己的电脑沦落到了数字黑社会中。同时，在管理者的压力下，软件开发者只顾得上不断推出最新功能，而可以减少软件漏洞的 bug 修复和代码重构只能一拖再拖。开发团队的管理者为了满足预算和最后期限，只能把成堆的技术欠债丢在脑后，任其不断逼近。

高管们总是迫切要求自己的团队第一个向市场提供新的数字解决方案，至于这些方案是否真的契合大众需求或会不会遭到黑客攻击则主要靠祈祷。项目经理会把富余的时间、经费和项目范围完全压榨掉，以满足项目截止期、预算和需求。测试被削减和推迟，以加速产品交付。威胁模型分析被斥为不切实际和偏执：哪个脑袋正常的人会偷我们的东西？

网络空间没有建设标准；安全性和韧性对于消费者，甚至电子商务供应商来说都是看不见的特点。随着投入网络的经济价值越来越大，有组织的破坏者、犯罪分子和恐怖分子正在利用那些防范薄弱的地方，例如索尼电影娱乐公司。虽然没有公布全部的事实，但从索尼的经验中互联网公司可以学到几个重要的教训。

### 网络攻击进化

虽然一些人把索尼事件归类为“网络破坏”，但它和金沙赌场在 2014 年遭遇的黑客攻击可能标志着网络攻击进化到了一个新阶段。在 2012 年和 2013 年，类似的恶意软件感染了石油巨头沙特阿美公司、几家韩国银行和媒体公司的 30000 台计算机。在金沙赌场的案例中，

行凶者想要惩罚这家公司，或者更准确地说，这家公司的首席执行官和头号股东亿万富翁谢尔登·埃德森 (Sheldon Adelson)。虽然要证实他们的猜测需要一些时间，但公

司高管们几乎立刻怀疑攻击是来自伊朗的。<sup>1</sup>

30 年前我们经历的那些按现在的标准看就是黑客们可笑和古怪的玩闹。例如，在政府网站上发一些幼稚的图片，烦人但是无关紧要的计算机病毒，还有大学生在计算机实验室里写出来的网络蠕虫。

当我们越来越多地把个人和公司的财富放到互联网上时，不法分子比较了网络犯罪和现实世界中犯罪的效益成本比，狡猾的骗子们开始制造间谍软件、钓鱼攻击和银行木马。然后，这类活动通过加密和勒索软件升级为了网络敲诈。最近，我们发现大规模的恶意软件制作活动已经进行了多年了。这些恶意软件包括大量收集旅行数据的间谍软件、Regin 恶意软件和有针对性地破坏离心机的 Stuxnet 病毒等。

最新的攻击方式则是焦土化的服务器和台式机数据删除，数 TB 敏感数据的渗漏，在科技新闻网站 ReCode 获得的一份索尼内部备忘录中，索尼雇来调查这次攻击的凯文·曼迪尔 (Kevin Mandia) 称，

事实上，这次攻击的范围与我们之前应对的任何攻击都不相同，它不但要破坏公司财产，还要把机密信息发布给公众。最重要的是，这是一次有组织的、前所未有的、计划周详的犯罪。对于这样的犯罪，无论是索尼电影娱乐公司还是其他公司都不可能完全抵御。<sup>2</sup>

索尼事件表明，网络攻击的数字本质使得攻击者可以用假情报和虚假线索布下烟幕，令人难以确定谁是罪魁祸首。几乎没有证据，更不用说决定性证据，可

以把犯罪者和这次攻击联系起来。很难排除这样的可能：这些攻击可能是商业竞争对手精心策划并在巧妙伪装下执行的。让一家公司遭受胁迫可以令它成为诱人的收购目标。

## 虚拟安全的成本与收益

IT 安全风险分析面对的巨大挑战是与生俱来的。我们很容易计算人工、流程和技术的成本，但只有在坏事发生了并且缺少安全控制，或安全控制不足以阻止或限制漏洞的影响时，安全收益才是可以计算的。而索尼的 IT 安全管理者们——用科技记者和评论员克什米尔·希尔 (Kashmir Hill) 的话说——“看来没有考虑过难以立刻计算的数据泄露成本，例如对公司声誉的打击，让雇员失去对公司的信任，或者詹姆斯·弗兰科可能不愿意让全世界知道，为他自己开车来电影拍摄现场，剧组就要支付 6000 美元。”<sup>3</sup>

还有一些其他指标可以反映索尼的网络安全状况，据希尔说，

信息安全团队相对来说很小。泄露的文件中有一份公司的花名册，上面列出了索尼电影娱乐公司的近 7000 名员工，其中只有 11 人属于一个头重脚轻的信息安全团队。团队中有 3 个信息安全分析师，而他们 3 人之上有 3 个经理、3 个董事、1 个执行董事还有 1 个高级副总裁。

3 个安全工程师不足以保障一个企业的安全。大型组织中变动的组成部分太多，这么小的团队无法完成自己的任务。而背后有 8 个管理人员盯着，显然无益于提高员工的工作效率和士气。

我明白，像许多其他公司一样，索

尼的网络架构是扁平的。这意味着，一旦外围网络被突破，没有内部防火墙或其他障壁能做出抵抗。公司员工、承包商、志愿者、合作伙伴、供应商、来宾，还有攻击者很大程度上可以自如地在网上移动。在塔吉特的数据泄露案例中，攻击者先攻陷了塔吉特的供应商的账号，破解了塔吉特的活动目录管理员密码，最终感染了销售终端系统。<sup>4</sup>

为了阻止这类攻击，服务器应该由商业部门专用。而且，在服务器之间，还有服务器和桌面计算机局域网之间都应该有防火墙分隔。只在互联网出口处设一个防火墙是不够的，它只是防御的第一道防线。只要有一个好奇的用户打开了看似来自 HR 的工资电子表格，攻击者就可以轻松地在防火墙内获得立足之地。

## 纵深防御

由于内网可能被攻破，企业、小商户，甚至家庭用户都应采取纵深防御战略。如果布拉德利·曼宁 (Bradley Manning) 和爱德华·斯诺登 (Edward Snowden) 教会了我们什么，那是要把网络划分为飞地或信任区域。分段网络可以阻止或减缓攻击者或内部人士从他们所在区域之外获取数据。从授权用户那里窃取到认证信息的外人跟怀有恶意的内鬼之间是没有区别的（除了你不能解雇前者）。

分段的网络架构<sup>5</sup>依赖于稳健的身份和访问管理 (IDAM) 解决方案，以确保得到认证和授权的用户可以使用数据。使用 IDAM 加入或移除员工并管理他们的访问权限和特权可以消除大量孤立的应用账户，这些账号有可能被心怀不满的前雇员利用或分享给攻击者。同样，也能制止软件开发者自己编写用户和密码管理方案，这类自己编写的软件很容易被

网络攻击者利用，成为攻陷系统的跳板。

索尼事件也说明了加密的重要性。

希尔指出，“一个因数据泄露而暴露出的明显问题是，索尼电影公司网络上的敏感文件没有在内部加密或受密码保护。”<sup>3</sup>所有关键数据在传输或者存储时都应该被加密。

因为企业应用开发成本高昂，默认的措施就应该是加密，而不是让你的知识资产暴露于数据渗漏和窃取的危险之下。如果值得实行自动化，那肯定也值得采取保护措施。因为内部网络设备、工作站和应用程序都可能被攻击，数据在传输时，从头到尾必须始终处于加密保护下。此外，认证管理必须集中在 IDAM 中，而不能分散在应用程序代码和配置文件里。

监测网络需要必备的工具，安全工程师应该使用 Splunk、LogLogic 和 Guardium 等工具确定数据流的基线。然后监测网络区块之间的数据流，还有通向外部的数据流。从金融区域流向设备管理区域的 1GB 数据可能标志着犯罪正在进行。如果数据是加密的，而且秘钥安全地保存在 IDAM 中，那么在追查犯罪者期间的损失是可以控制的。

必须用应用程序白名单代替病毒特征库。多态病毒意味着特征库是愚蠢的策略，相反，应该阻止用户运行任何未经批准的程序。“应用程序白名单早在十年前就已经出现了，但直到最近几年，这类产品才趋于成熟，变得易于管理”（见 <https://isc.sans.edu/forums/diary/Defensible+network+architecture/19141/>）。现在，微软已经开始为企业用户提供 AppLocker<sup>6</sup>，我们再也不能以此为借口了。

把软件保障构建到你的软件开发生命周期中。无论大小，几乎所有组织都

会发现自己与软件业有所牵连。从修改商业软件包到完全由自己开发应用软件，企业在不断增加存在漏洞的软件。一些开发者宣称防火墙可以阻挡大多数攻击，因此修正代码是没有必要的。但由于钓鱼攻击、有漏洞的网络应用程序和心怀不满的内部人士的存在，即使是内部系统也会成为攻击的目标。为了应对这些威胁，开发团队应该利用攻击树、威胁模型、滥用用例和安全设计模式。他们可以采用正规方法，用数学成熟的技术来确保软件的安全性。最后，他们的经理应该看到，开发团队正在采取适当的措施防止软件出现漏洞。只有防火墙是不够的。



益于索尼那些垃圾话邮件在晚间新闻上的亮相，世界各地的企业高管应该都在问他们的网络安全管理人员，“这样的事能在我们这发生吗？”当涉及到网络安全的投资时，影响巨大的公开羞辱胜过一大堆最佳成本效益分析。如果你的领导问，“索尼的数据泄露能在这发生吗？”你会如何回答呢？IT

“Sony Says Hack Attack Is Unprecedented,” Recode, 7 Dec. 2014; <http://recode.net/2014/12/07/sony-describes-hack-attack-as-unprecedented/>.

3. K. Hill, “SONY Picture Hack Was a Long Time Coming, Say Former Employees,” Fusion, 4 Dec. 2014; <http://fusion.net/story/31469/sony-picture-shock-was-a-long-time-coming-say-former-employees/>.

4. T. Olavsrud, “11 Steps Attackers Took to Crack Target,” CIO Magazine, 2 Sept. 2014; [www.cio.com/article/2600345/security/0/11-steps-attackers-took-to-crack-target.html](http://www.cio.com/article/2600345/security/0/11-steps-attackers-took-to-crack-target.html).

5. R. Bejtlich, *The Practice of Network Security Monitoring*, No Starch Press, 2013.

6. M. Graven, “Lock Down Unauthorized Applications with the Built-In AppLocker Tool,” Microsoft TechNet Magazine, <https://technet.microsoft.com/en-us/magazine/dd547414.aspx>.

## 参考文献

1. B. Elgin and M. Riley, “Now at the Sands Casino: An Iranian Hacker in Every Server,” Business Week, 11 Dec. 2014; [computer.org/ITPro\\_57\\_hit-sheldon-adelsons-sands-casino-in-las-vegas](http://www.businessweek.com/articles/2014-12-11/iranian-hackers-polymorphic-viruses-can-mean-that-signatures-are-a-sucker-s-game-instead-prevent-users-from-running-any-software-that-isn-t-approved-computer-org/ITPro_57_hit-sheldon-adelsons-sands-casino-in-las-vegas).

2. A. Hesseldahl and D. Chmielewski,

沃尔特·豪泽（Walter Houser）应用安全工程师，正支持一家大型政府组织。他曾领导与多个联邦机构合作的信息保障项目。他曾作为 IT 政策官、网站管理员、应用开发经理，以及企业架构师。联系方式：houser@owasp.org。



## 虚拟反乌托邦

托尼·施利斯基 (Tawny Schlieski), 英特尔 (Intel)

足够先进的技术无异于邪恶

——改编自亚瑟·C·克拉克的名言

# 过

去的 7 年中，我一直在好莱坞与电影制作者们一起工作，探索数字创作那极富想象空间的深厚潜力。这些年来，该领域的进步是惊人的。6 年前詹姆斯卡梅隆在拍摄《阿凡达》时率先使用的虚拟制作工具，正在重塑电影行业工作的本质。曾被人们坚决地从日常创作流程中摒弃的数字创作，现在已经成了导演、设计师和演员实际工作流程的一部分。整个行业的艺术家现在深入到完全由计算机渲染而成的虚拟环境中，规划和想象他们的电影。演员以虚拟角色的形式表演，最终真正地从他们物理形态的限制中解放出来。导演在拍摄现场可以在摄制过程中看到他们的故事的物理和数字元素，而人们熟悉的那句“我们会在后期制作时修正这个问题”也变成了“让我们再试一次。”

## 数字世界

充分发挥作用的数字资源和反应灵敏、规则驱动的数字环境，正在颠覆我们设计和创造一切事物——从建筑物和飞机到城市和

工地的方法。用于创造数字世界和角色的工具正不断进化，以更好地满足使用它们的艺术家的需求。这些工具，如梦工厂的阿波罗 ([www.screendaily.com/home/blogs/dreamworks-animation-unveils-apollo/5072353.article](http://www.screendaily.com/home/blogs/dreamworks-animation-unveils-apollo/5072353.article))

或迪士尼的摄影捕捉系统 ([www.awn.com/vfxworld/all-worlds-virtual-stage-disney-s-new-camera-capture-system](http://www.awn.com/vfxworld/all-worlds-virtual-stage-disney-s-new-camera-capture-system)) 已经具备了可以流畅工作，并与使用者的想象力保持步调一致的计算能力。英特尔和美国南加州大学的研究人员已经开始利用这些工具，不断发展虚拟现实 (VR) 和增强现实 (AR) 技术，建立数字化的世界，并探索新的叙事体验形式 (见图 1)。随着这些工具的进步，学习并掌握它们的艺术家的数量正在以几何级数增长。

现代电影摄影机是数字创作的一个例子。它的神奇之处不在于它的现实，而是超现实。有了它，艺术家，科学家和宣传人员就能方便地以真正新颖的方式表达自己的想法：一个人可以变成龙。我们可以在数秒内看遍一个城市 10 年间的发展历程。无需动工就可以测试新型建材。士兵们可以训练实战能力，而不用冒生命危险。

## 谁害怕虚拟现实？

那么，为什么围绕虚拟现实 (VR) 会出现这么多争议？除

## SCIENCE FICTION PROTOTYPING

除了怀疑 VR 技术的可行性外，还有人担心这种技术的成功所带来的后果。许多人担心的是 VR 与人工或机器智能的结合。奇点的概念——机器智能给地球生命带来威胁——往往与虚拟和增强现实技术的出现联系在一起（[www.33rdsquare.com.com/2013/07/nick-bostrom](http://www.33rdsquare.com.com/2013/07/nick-bostrom)）。这样的观念流传广泛，不仅局限于在 YouTube 上看过尼克·博斯特伦（Nick Bostrom）名为“人类的末日”的 TED 演讲（[www.youtube.com/watch?v=P0Nf3TcMiHo](http://www.youtube.com/watch?v=P0Nf3TcMiHo)）的 70000 人。

几十年来，艺术家们把 AI 与 VR 纠缠在一起，用类似于伊卡洛斯和弗兰肯斯坦博士的幻想故事描绘 VR 技术。这些故事警告我们，人类野心和抱负是多么傲慢自大，试图超越自然或神圣的界限会带来怎样的危险：伊卡洛斯飞得太靠近太阳，弗兰肯斯坦创造了生命。与此类似，在沃卓斯基（Wachowsky）姐弟 1999 年的电影《黑客帝国》中，疯狂的 AI 机器奴役着人类，而在尼尔·斯蒂芬森（Neal Stephenson）1992 年的小说《雪崩》中，虚拟实境（Metaverse）诱惑人类开启它，沉迷于它并脱离现实。这也难怪我们不相信这些工具。我不知道你怎么想，但我真的不想成为人体电池。

**如**果我们摆脱那些存在于想象和隐喻中的 VR 形象，我们将看到一个沿着与斯蒂芬森和沃卓斯基姐弟的梦想完全不同的道路发展的创造性工具。事实上，这些新工具不但不会让人类边缘化，还能培育出全新一代的创作者。动作捕捉没有取代演员——它赋予他们此前不敢想象的饰演角色的能力。数字世界没有取代我们的现实——它们使艺术家能以全新的方



图 1.1 在 2014 年的国际消费电子展（CES）上，英特尔和南加州大学的利维坦（Leviathan）项目（名字源于史考特·韦斯特费德（Scott Westerfeld）的同名小说），带领参观者步入了一个鲸在人们周围飞翔的虚拟世界

式可视化地呈现科学和幻想。

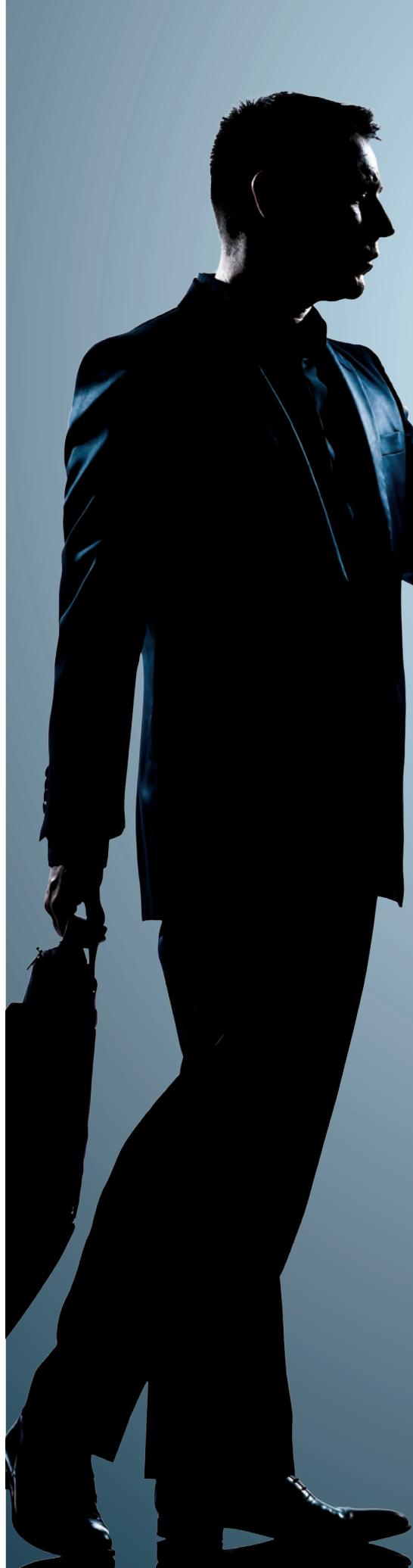
在拍摄 2014 年上映的电影《星际穿越》时，导演克里斯·诺兰（Chris Nolan）请天体物理学家基普·索恩（Kip Thorne）提供指导，帮助 Double Negative 特效公司设计黑洞视觉效果（[www.wired.com/2014/10/astrophysics-interstellar-black-hole](http://www.wired.com/2014/10/astrophysics-interstellar-black-hole)）。经过一年的工作，重写了渲染引擎，并制造了数百 TB 的数据后，他们创造出了“准确程度前所未有的模拟黑洞”，

让艺术家和科学家都大为惊叹。Double Negative 投入在黑洞上的计算能力没有凌驾于最终的效果：它只是艺术家和科学家共享的一种工具，像之前那些工具一样，彻底地重塑我们看待世界的方式。■

**托尼·施利斯基（Tawny Schlieski）**是英特尔公司的研究科学家，她的联系方式是 [tawny.schlieski@intel.com](mailto:tawny.schlieski@intel.com)。

**IEEE Software** 寻找可以吸引专业和非专业读者的实用可读文章。这本杂志的目标是将可靠的信息传递给软件开发者和管理者，帮助他们站在技术浪潮之巅。投稿必须为原创，不能超过 4700 个词，每张表格和图片不能超过两个百词。

**投稿指南：**  
[www.computer.org/software/author.htm](http://www.computer.org/software/author.htm)  
更多细节：[software@computer.org](mailto:software@computer.org)  
[www.computer.org/software](http://www.computer.org/software)



# 会议就在你的手中

IEEE计算机协会的会议发布服务（CPS）现在可以提供组织会议的移动应用了！让会议的日程、会议信息和论文列表在你的与会者手中的设备上显示。

会议的移动应用可在[安卓](#)设备、iPhone、iPad和Kindle Fire上运行。



欲知更多信息，请联系[cps@computer.org](mailto:cps@computer.org)





**史蒂文·M·贝洛维**  
Steven M. Bellovin

哥伦比亚大学  
Columbia University

# 加密到底应该是什么？

**关**于密码学，我们只知道一件事：它很难。事实上，它在所有层面上都很困难：无论是原始的加密原理，加密协议，加密的实现，还是使用规则，都非常非常困难。而如果犯错，就会被攻击者加以利用。很多问题已经在技术社区中引起了广泛的关注。参加国家标准技术研究所加密或哈希算法竞赛的提交必须包含安全分析，要能抵挡已知的攻击类型，例如差分密码分析。

设计糟糕的协议必然存在风险。不过，大部分攻击规模不大，这些攻击一般要求大量的截取数据——这本身对很多攻击者来说就是很大的挑战，还需要技术高超的攻击者。然而，有一类问题并没有在可用性（usability）社区外引起多少注意，这类问题就是用户犯下的错误。忽略这个问题后果严重，用户的错误过去和现在都能造成严重的威胁，因为他们也会在不知道的情况下发送明文信息。

这并不意味着没有警告。长期以来，密码会因为工作人员产生的信息没有加入足够的空值，或充分使用密码文本的同义替代而被破解。德国在使用二战时期的恩尼格玛密码机时犯了错，帮助了英国的密码破译人员。更近一些，惠滕（Whitten）和泰加尔（Tygar）在经典的《为什么约翰尼加不了密？》（Why Johnny Can't Encrypt）是一个明确的警告，但是后续的研究要么面向可用性社区（也是这一观点的鼓吹者），要么至少像专注于可用性问题那样专注于协议性问题。这么做是错误的。可用性失败是钓鱼攻击和意外明文邮件的主要技术原因，而且在还时常导致网页公钥基础设施出问题。

让我们更仔细地研究一下被加密保护的电子邮件：它应该是什么样的？发送者是不是必须要求加密？用户能否在没有密钥的设备上接受信息？如

果默认开启加密，用户是否应该有权关闭加密？如果已知某些收信人有能力或无法进行加密，或者不知道收信人是否有这种能力，应该怎么办？怎么区分这几种收信人？我们知道用户并不会留意非常细小的提示，比如一个带锁的图标；我们也知道服务商因为自身的原因很喜欢大幅度改变UI——可以比较一下iOS6和iOS7——所以提示图标也不管用。

收信人也有相似的问题。他们如何被清楚地告知某条信息在加密的条件下被接收？如何显示某个数字签名是否存在或缺失？如何显示加密验证的发件人，而不是显示一行“寄件人：”？如果他们有分歧怎么办？如果加密验证的发件人和“寄件人：”那一栏显示的发件人不符，怎么办？如果我们可以在一个晚上升级所有电子邮件的程序系统，这些问题就会相对简单，但很明显我们没法这么做。

密钥处理本身也存在挑战。我怎么才能安全地决定他人的公共密钥？要怎么做我才不需要担心甚至知道公共密钥？他们如何才能在不理解诸如证书和PKI等重要概念的情况下，获得我的密钥？在现今操作系统糟糕的安全性的状况下，我如何保护自己的私钥？如何处理例外情况，例如密钥吊销？

这些问题全都没有简单的答案。我们甚至还不清楚有些问题是不是会有答案。我们不知道，除非整个安全社区共同关注一个看起来相对简单的问题：加密到底应该是什么？

（本文内容来自 Security & Privacy）

SECURITY & PRIVACY

**史蒂文·M·贝洛维**（Steven M. Bellovin）是哥伦比亚大学的计算机科学教授。他的网也是：<https://www.cs.columbia.edu/~smb/>。



## 全新微信号上线

科研圈

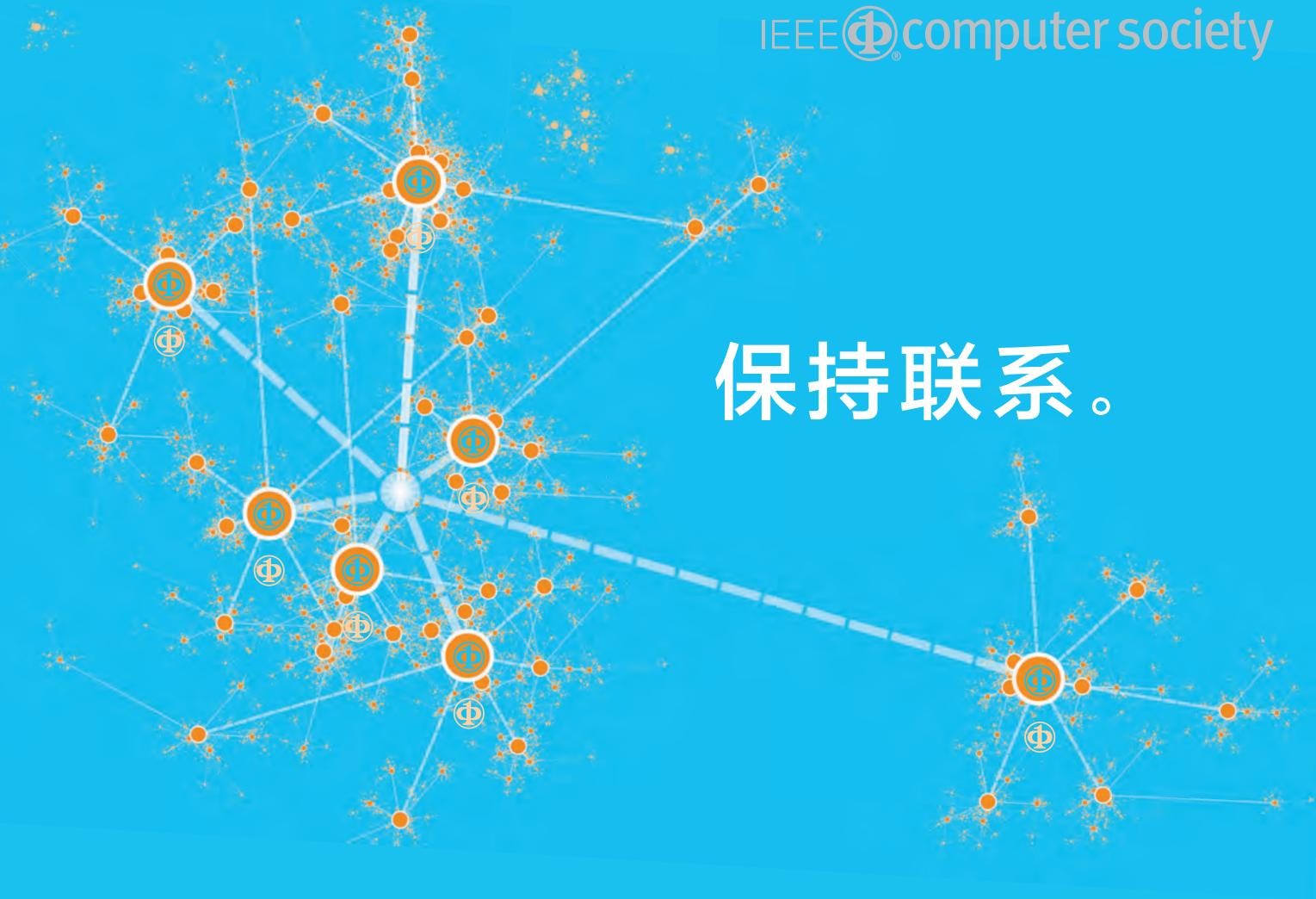
# 连接科研圈 服务实验室

汇集顶级学术期刊中文摘要，解读一线研究成果  
专访全球知名实验室与科学家，分享科研方法与理念  
推送精彩讲座与会议预告，科学大事不漏报

微信号: keyanquan



环球科学出品



# 保持联系。

无论你在哪里，都能紧随IEEE计算机协会的脚步。

在Twitter、Facebook、Linkedin和YouTube上关注我们。



@ComputerSociety, @ComputingNow



[facebook.com/IEEEComputerSociety](http://facebook.com/IEEEComputerSociety)  
[facebook.com/ComputingNow](http://facebook.com/ComputingNow)



IEEE Computer Society, Computing Now



[youtube.com/ieeecomputersociety](http://youtube.com/ieeecomputersociety)

绿色印刷 保护环境 爱护健康

亲爱的读者朋友：

本书已入选“北京市绿色印刷工程——优秀出版物绿色印刷示范项目”。它采用绿色印刷标准印制，在封底印有“绿色印刷产品”标志。

按照国家环境标准（HJ2503-2011）《环境标志产品技术要求 印刷 第一部分：平版印刷》，本书选用环保型纸张、油墨、胶水等原辅材料，生产过程注重节能减排，印刷产品符合人体健康要求。

选择绿色印刷图书，畅享环保健康阅读！



北京市绿色印刷工程

绿色印刷产品