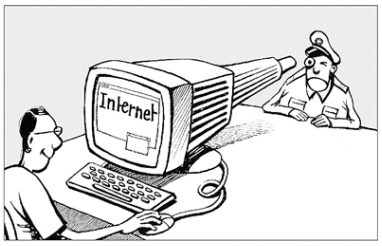
棱镜计划与贝叶斯公式*

自 2001 年“9 • 11”事件发生之后，美国情报机构加大了监视范围，旨在避免恐怖袭击事件的再次出现。2007 年小布什当政时期，美国国家安全局和联邦调查局[1]启动一个代号为“棱镜”的秘密项目，布什政府开始了所谓的恐怖分子监视计划（Terrorist Surveillance Program），参与行动的政府部门可以在无法院批准的情况下秘密监控美国境内的电话和电子邮件。

2013 年 6 月，爱德华 • 约瑟夫 • 斯诺登（Edward Joseph Snowden，1983 年 6 月 21 日－）将“棱镜计划”的秘密文档披露给《卫报》和《华盛顿邮报》。他“愿意牺牲掉这一切（工作、收入和女朋友）（把真相告诉世人），因为美国政府利用他们正在秘密建造的这一个庞大监视机器摧毁隐私、互联网自由和世界各地人们的基本自由的行为让他良心不安”。泄露文件显示，棱镜计划能够对即时通信和既存资料进行深度的监听。[2]

棱镜计划标志



对用户深度监听

棱镜是如何工作的呢？查阅网上一些资料，发现棱镜的工作原理并不复杂，它对用户通信信息进行监听，一旦出现敏感信息，它就会向上级反映，并追踪信息来源。举例来说，一个从伊朗 IP 地址登录的用户，使用 Google 搜索一些信息，或 MSN 发送一条信息，思科[3]的设备就会把通讯信息记录下来，而后直接分析通讯内容。如果该用户提到“真主”、“阿拉”、“爆炸”这样的词，安全机构则向谷歌或微软等公司提出请求，要求其提供该用户的相关资料并进一步调查。

*本文作者高磊、宋培培。

[1]国安局（NSA）听命于国防部长，也就是说属于军方，负责全球监听和技术侦查，相当于军队的喉舌；中情局（CIA）直接听命于总统，也就是说属于政府，除了协调全美情报机构的工作外，还负责执行一些秘密行动，也就是我们常说的间谍和反间谍活动，相当于政府的喉舌 + 手脚；联邦调查局（FBI）听命于总统，隶属于司法部，虽然也属于政府部门，不过级别就比 CIA 低了半级，主要是针对国内进行执法调查的，尤其是对一些重案和需要跨州协调的联邦案件开展调查，当然也包括国内的反间谍工作，因为不是间谍部门，所以 FBI 的工作都是公开的，没有 CIA 那么神秘，也不搞暗杀之类的活动。

[2]更多请参阅 爱德华 • 斯诺登/棱镜计划。

[3]思科系统公司, 大型互联网设备提供商，产品有路由器、交换机、光线平台等。

思科路由器把持着中国骨干网络的超级核心节点

棱镜计划曝光后，许多政治人物强烈不满。据传，美国国安局已监听德国总理默克尔长达 11 年，默克尔强烈谴责间谍行动，并要求立即停止美国国家安全局的监听行为。巴西总统罗塞夫对被监听极为愤怒，要求美国公开道歉。除此之外，平常百姓家也深受其影响。卡塔拉诺 (Michele Catalano) 家住纽约纳苏县，因为自己在网上搜索了一些敏感词，招致“反恐联合工作组”特工前去她家搜查。原来，他们家想买一个高压锅，丈夫用 google 搜索了“高压锅”，不久她用 google 搜索了“背包”。时值波士顿爆炸案[4]发生不久，这些词汇较为敏感，特工就找上门来搜查。

网上查高压锅资料，惹 FBI 特工找上门

这则新闻引起了不小的轰动，央视《新闻 30 分》以一种调侃的口吻报导了这则新闻[5]。许多人并不认同央视的这种态度，他们认为这则新闻恰恰体现了美国情报人员的专业素质：对网络信息的搜集分析以及快速反应能力。不过，这种方法真的靠谱吗？其准确性到底有多高？为什么会闹出“高压锅”的笑话？接下来，我们从统计的角度进行理性分析。

如果恐怖分子发动恐怖袭击，与同伴之间的通讯很可能包含一些敏感词汇，如“真主”、“圣战”等。用条件概率来描述上述事实：

$$P(+|\text{坏人}) = 0.99$$

其中，“+”表示特殊词汇，上式表示恐怖分子使用“+”的可能性高达 99%。

问题是，如果检测到某人使用“+”词汇，我们有多大的把握判断这个人就是恐怖分子：

$$P(\text{坏人}|+) = ??.$$

很自然的想法是：既然恐怖分子使用“+”词汇的可能性如此之高，如果某人使用了“+”词汇，那么某人是恐怖分子的可能性也应很高。然而，这种逻辑是忽略了很重要的事实：人群中恐怖分子极其稀少。

[4]波士顿爆炸案中，恐怖分子将火药、电话板、钉子、弹头与滚珠塞进高压锅，并把高压锅装进一个背包。

[5]“网搜两词汇警察找上门”(2013/8/3《央视新闻 30 分》)。

我想做好人

补充两组合理的信息。其一，坏人（特别是恐怖分子）是极为罕见的，假设其概率为 50 万分之一[6]：

$$P(\text{坏人}) = 1/500,000.$$

其二，好人使用“炸弹”“砍刀”这类“+”词汇频率低于 10%：

$$P(+|\text{好人}) = 0.10.$$

结合以上信息，运用贝叶斯公式推导使用特殊词汇的某人为恐怖分子的概率：

$$\begin{aligned} P(\text{坏人}|+) &= \frac{P(+|\text{坏人}) * P(\text{坏人})}{P(+|\text{坏人}) * (\text{坏人}) + P(+|\text{好人}) * P(\text{好人})} \\ &= \frac{0.99 \times \frac{1}{500000}}{0.99 \times \frac{1}{500000} + 0.10 \times \frac{499999}{500000}} \\ &= 0.0000198 \end{aligned}$$

也就是说，即使一个人搜索了“+”敏感词汇，他是恐怖分子的概率仍非常低, 仅为 0.0000198，换句话说，使用敏感词汇的 50000 个人中，只有一人是恐怖分子。

故而，且不说棱镜计划毫不顾忌公民隐私招致非议，即便从技术角度考虑，以该计划为代表的一大批监听项目，也未发挥它们应有的作用。这值得安全人员思考。

参考资料

- How likely is the NSA PRISM program to catch a terrorist?
- 棱镜系统是怎么工作的？
- 维基百科：棱镜计划

[6]美国国家反恐中心公布 2011 全球恐袭报告提到，2011 年全球有 13288 人死于恐怖袭击，以全球人口 70 亿计，死于恐袭的概率仅有 0.0002%, 换句话说，50 万人中约有一人死于恐怖袭击。有人比喻，死于恐怖袭击的概率低于在家被家具砸死的概率。