SecurityScorecard

Cyberworld

深入探究评分机制

作者：数据科学博士兼 SecurityScorecard 副总裁 Bob Sohval

SecurityScorecard

Cyberworld

目录

SecurityScorecard

Cyberworld

SecurityScorecard采用“由外而内”的方法，非侵入性地评估企业的安全状况。

这种方法使SecurityScorecard能够大规模运行，每天提供和更新全球一百多万个企业的网络安全评级。

网络安全评级

互联网的兴起及其在电子商务、商业运营、通信和社交媒体中的全球作用，既创造了机遇，也带来了网络风险。虽然它可以促进经济增长，加速资讯和思想的传播，但常用软件产品和服务中存在漏洞，以及未严格遵守建议的安全措施，可能会使企业遭受恶意行为者（包括个人和民族国家）的重大财务和声誉损害。

网络安全评级提供了一种方法，可以客观地监察企业的网络安全态势，并衡量其网络安全态势随着时间推移的变化：是改善还是恶化。网络安全评级对于供应商风险管理计划、确定网络保险的风险保费、信用承保和金融交易决策、并购尽职调查信息、高管层报告以及自我监察都非常有用。网络安全评级及其所基于的大量信息也有助于评估对网络安全风险标准的合规性。

评分意味着什么?

SecurityScorecard通过总评分传达对企业网络安全态势的详细分析,总评分以易于理解的字母评级表示:从 A (90-100) 到 F (<60)。总评分直接反映了SecurityScorecard使用问题类型权重在企业面向互联网的资产上发现的所有安全问题。

网络安全评级可与金融信用评级相比较。例如,信用评级较差会导致违规概率增加,网络安全评级较差会导致数据泄露或其他不利的网络事件发生概率增加。

使用统计分析验证SecurityScorecard评分表明:与评级为 A 的企业相比,评级为 F 的企业发生数据泄露的可能性高出 13.8 倍。

因素评分

SecurityScorecard 可计算并提供 10 种不同因素评分的详细报告。因素评分将网络风险的不同方面分组并多维度地进行描述。它们可让安全团队识别易受攻击的区域,并将补救措施集中在最能产生影响的地方。

分数范围是0-100。问题类型根据相对的泄露风险进行加权。问题类型权重是唯一影响总评分的权重。这使得评分计算过程清晰易懂。

单个因素评分是根据与该因素相关的安全问题、或调查结果的严重程度和发现的安全问题数量计算得出的。

因素评分为100表示:未检测到该因素的网络安全问题。

评级	评分
A	>90
B	80-89
C	70-79
D	60-69
F	<60

SecurityScorecard

Cyberworld

SecurityScorecard 的 10 个风险因素组

1. 应用程序安全
2. Cubit 评分
3. DNS 健康状况
4. 端点安全
5. 黑客情报
6. 信息泄露
7. IP 信誉
8. 网络安全
9. 补丁修复
10. 社会工程

网络安全数据

SecurityScorecard监察数百种不同的网络安全数据，根据定义的问题子集问题计算分数。每个问题都与十个风险因素组之一相关联，并根据其与违规可能性的密切相关性分配一个反映其严重程度的权重。信息性和积极性问题（反映良好的安全实践）被捕获并呈现给用户以提高意识，不会影响评分。

请注意：

- 随着SecurityScorecard不断改进和优化评分算法，严重程度级别可能会发生变化。这些变化将作为SecurityScorecard季度评分重新校准的一部分进行，版本号也将相应更新。
- 在SecurityScorecard平台上可以找到每种问题类型的详细描述、风险和建议。

数据处理的工作流程

生成有意义的网络安全评级包括四个不同的处理阶段：数据采集、归因引擎、网络分析和评分引擎。

数据采集

- IPv4 扫描
- 恶意软件沉洞
- DNS 数据
- 外部数据源

归因引擎

- RIR、DNS、SSL 数据
- 域名发现
- 子域名
- IP 域名配对

网络分析

- 研究新兴威胁
- 公共漏洞和暴露（CVE）
- 机器学习

评分引擎

- 数字足迹
- 规模标准化
- 因素评分
- 总评分

数据采集

SecurityScorecard定期扫描整个IPv4网络空间，以识别易受攻击的数字资产。此外，SecurityScorecard通过遍布美洲、亚洲和欧洲的全球传感器网络，监察整个互联网上的数据。SecurityScorecard运营着世界上最大的沉洞和蜜罐网络之一，用于捕获恶意软件，并通过商业和开源情报来源进一步丰富数据集。SecurityScorecard通过大约40个第三方公共和商业数据源的外部数据来补充其数据采集。作为数据采集计划的一部分，SecurityScorecard每天大约采集1.5TB的数据。

归因引擎

大多数采集到的数据都与IP或相关域名相关联，然后必须根据其数字足迹，将其与企业进行匹配。由于互联网的动态特性，IP归因是一个具有挑战性的过程。互联网服务提供商(ISP)、云服务提供商 (CSP) 和内容分发网络 (CDN) 可以动态分配 IP 的网络块。这些可能每天甚至每小时都会发生变化。此外，由于互联网的分布式特性，DNS更新可能需要一些时间才能在整个网络中传播。

从根本上讲，归因是一个随机或概率过程，而不是一个确定性过程。这意味着，从实际角度来看，归因永远不可能100%准确。然而，有了高质量的数据源和先进的算法，错误率可以保持在相当低的水平。

SecurityScorecard使用在互联网规模上运行的自动化流程进行归因，并结合机器学习算法来优化准确性。

SecurityScorecard使用RIR、DNS、SSL和其他方式以及第三方数据源将IP归因于域名。每个数据源都有自身的置信度，因此会针对每个候选域名IP对汇总数据源，如果总体置信度令人满意，则接受该域名IP对。IP数字足迹每天更新。

除了IP归因，SecurityScorecard还运行域名发现程序，查找每个评分企业控制的相关域名和子域名。

对于每个评分卡，SecurityScorecard使用域名WHOIS服务以及被动DNS源生成相关域名列表。然后，使用统计技术和子字符串匹配处理该列表，以仅保留高置信度相关的域名。

根据独立专家的渗透测试，错误地将一个域名归因于某个企业的误报率通常低于 5% 。

SecurityScorecard使用内部系统执行子域名发现，该系统使用来自CommonCrawl、SSL认证以及多个商用数据源的数据。子域名被解析为 DNS A 记录并归父域名所有，因此误报率非常低。

根据一家安全公司的独立评估，域名归因的误报率低于 1%

SecurityScorecard

Cyberworld

网络分析

SecurityScorecard部署了一套由其威胁情报研究人员、数据科学家和软件工程师开发的分析工具，用于从原始输入数据中提取和获得关键见解。关键分析、工程和数据处理的示例包括：

- 对恶意软件家族进行逆向工程，以便识别不同的恶意软件类别，描述其行为和威胁级别。
- 通过检查横幅抓取返回的数字资产以及分析网站代码库、通信协议和SSL认证来识别CVE和其他漏洞。
- 应用机器学习算法来提高安全调查结果的质量和准确性，并提供关于安全态势的关键见解。

评分引擎

评分是一个基于企业的数字足迹和观察到的风险情报的确定性过程。SecurityScorecard的评分引擎每天为全球超过1,200万家企业发布和更新评分。

评分机制

在为企业网络安全提供公平且准确的评级时，一个独特的挑战是：如何正确地考虑企业规模的广泛差异？

较小的企业，例如“MomAndPop.com”，其数字足迹规模较小，仅有一个或几个IP，与运营着数亿个IP的大型企业相比，必然会发现较少的安全漏洞。

如果不对数字足迹的规模进行修正，较大的企业会比较小的企业有更多的安全漏洞，则会获得更差的网络安全评分。

规模标准化

为了消除因规模造成的评分偏差，SecurityScorecard开发了一种基于强大统计框架的原则性评分方法。无论企业规模如何，都能确保评分是公平的。

许多类型的安全问题会随着企业规模的扩大而增加。与较小的企业相比，较大的企业往往具有更大的“攻击面”。员工越多，意味着需要保护的设备就越多。服务器越多，意味着暴露端口的机会就越多，而这些端口应该位于防火墙后面。有些问题类型会随着IP数量而变化，或者随着相关域名数量或员工数量而变化。

如上所述，不同企业的数字足迹从单个IP到数亿个 IP 不等。此范围跨越了 8 个数量级以上，或 10 的 8 个倍数以上。在如此大的动态范围内进行有意义的测量，最佳的方法是使用对数刻度，其中每个增量对应 10 的倍数。

使用对数刻度来比较大动态范围的测量值的其他常见示例包括以下几种：

- 里氏震级，用于测量 9 级以上的地震。
- 分贝标度，用于测量 12 个数量级以上的声音振幅。
- pH 值，用于测量 14 个数量级以上的化学酸度。

规模标准化从散点图开始，以捕捉特定问题的发生次数如何随企业规模而变化。

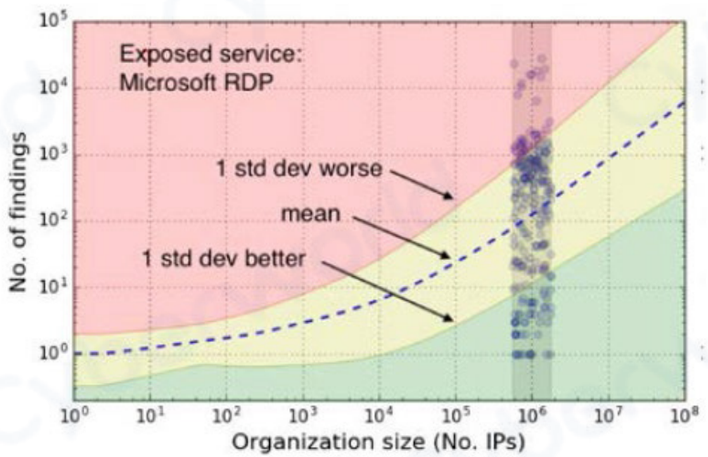
对于每个企业和每个安全问题，都会捕获问题类型的发生次数。**示例**是开放端口3389，对应于Microsoft的远程桌面协议。生成散点图，每家评分企业表示为双对数图上的一个点，其中Y轴是发现的安全问题数量的对数，X轴是企业规模IP数量的对数。典型的散点图包含数百万个数据点，提供大量统计质量，以实现更好的准确性和稳定性。

SecurityScorecard评分的企业数量庞大，目前已超过1,200万家。这有助于准确表达每种问题类型的发生次数随企业规模的分布，从而得到更准确的评分。

规模标准化使SecurityScorecard能够为用户提供评分依据。

在下一页的示例中，该企业有 3 个 DNS Open Resolver实例，这是一种DNS服务的错误配置，可被恶意行为者利用来发起DDoS攻击，可能导致业务中断和声誉受损。

根据SecurityScorecard对1,200万家企业的分析，在类似规模的企业中，只有12%存在此安全漏洞。此外，在类似规模且存在相同安全漏洞的企业中，发现此类安全漏洞的平均数量为2个，而该企业有3个，这比平均水平还要糟糕。



SecurityScorecard

Cyberworld

该企业与类似规模的企业进行比较

- 12%存在此类安全漏洞
- 88%不存在此类安全漏洞
- 平均数量为 2 个
- 该企业有 3 个

校准流程

SecurityScorecard会针对每种评分问题类型生成类似于上一页示例的散点图。然后，应用局部加权非参数拟合算法来表达企业规模的函数——预期问题数量的平均值（蓝色虚线）和标准差。

值得注意的是，问题数量与企业规模的关系是非线性的（蓝色虚线是曲线）。如果简单地假设问题数量与企业规模呈线性增长，就会引入严重错误，从而导致系统性扭曲和不正确的网络安全评分。

每种评分问题类型都执行此校准流程，使用在2个月的时间间隔内采集的数据来平滑统计波动。

校准流程使某企业能够与其他类似规模的企业进行公平的绩效比较。在上一页的示例散点图中，红色区域中的企业至少比平均值差 1 个标准差，而绿色区域中的企业至少比平均值好 1 个标准差。这种方法确保始终与其他类似规模的企业进行比较。

计算因素评分

上述校准流程能够为特定的企业和安全问题计算出可靠且稳定的统计估计值，对应于该企业在特定安全问题上高于或低于平均值的标准差。在统计学术语中，这被称为“z 分数”。

SecurityScorecard使用“修改后的z分数”，其中，如果未发现任何问题，则$z = 0$；如果发现问题数量等于具有相同数字足迹规模的企业的平均值，则 $z = 1$。在此框架中，$0 \leqslant z < 1$ 表示优于平均水平，而 $z > 1$ 表示低于平均水平。

计算原始总分

$$RTS_d = \sum_{i \in f} w_i \times z_{di}$$

在SecurityScorecard评分机制3.0版本中，不再使用因素评分来计算总分。SecurityScorecard通过将与问题发现相关的所有z分数乘以其权重或严重程度（低、中、高、严重）相加来计算原始总分（RTS）。

SecurityScorecard使用机器学习根据其与违规可能性的相关性来计算权重：相关性越大，严重程度越高。

计算总分

$$TS_d = MAX - \frac{MAX - TS_0}{\mu(x_d)} \times RTS_d$$

计算原始总分后，SecurityScorecard会根据问题发现计数的预期值对其进行缩放。SecurityScorecard希望通过将企业与其他具有类似数字足迹规模的企业进行比较来公平地评分。

信息性和积极性问题不影响评分。

违规惩罚

企业的数据泄露是安全入侵的外部证据，反映出风险增加。为了反映这种风险，在披露违规行为后，其评分会降低10%。惩罚对评分的负面影响在30 天内逐渐减小至零。

右侧的评分历史图展示了6月初发生的数据泄露的影响。违规惩罚将评分从90降至81，降幅为10%。惩罚对评分的影响在接下来的30天内逐渐减小，然后在7月初不再影响评分。

一段时间内的总评分

保持评分框架的最新状态

SecurityScorecard竭尽全力创建和维护有意义、准确且相关的网络安全评级。

由于网络威胁随着新威胁的出现以及新对策和最佳实践的发展而不断演变，就像军备竞赛一样，SecurityScorecard持续监察威胁形势，并评估新的数据源和新的分析方法，以更好地反映网络安全风险。

校准频率

SecurityScorecard会定期每月重新校准其评分算法。包括FICO、S&P和Moody’s在内的信用评级机构也会定期重新校准其评分算法。与网络安全风险评级相比，金融风险评级的标准相对稳定，因此校准频率较低。

保持定期的评分更新频率，使SecurityScorecard能够在动态威胁环境中保持公平的网络安全风险评级，并根据需要引入反映新风险指标的新问题类型，从而让用户及其生态系统更好地了解情况。

行业比较

上述校准和评分流程适用于平台上的所有企业。这种方法确保了庞大的统计质量，能可靠地测量和评估 1,200 多万家企业的安全态势。

每个评分企业都会被分配一个行业标签，以便于在行业内和跨行业进行比较。SecurityScorecard可以轻松地将各个企业的总评分和因素评分与同一行业内的其他企业进行对比。无论是在某个时间点，还是在长达 12 个月的时间段内分析趋势。

全局校准和评分还可以比较不同行业部门的整体安全态势，这对于网络保险承保和国家级的网络风险评估非常有用。

行业类别

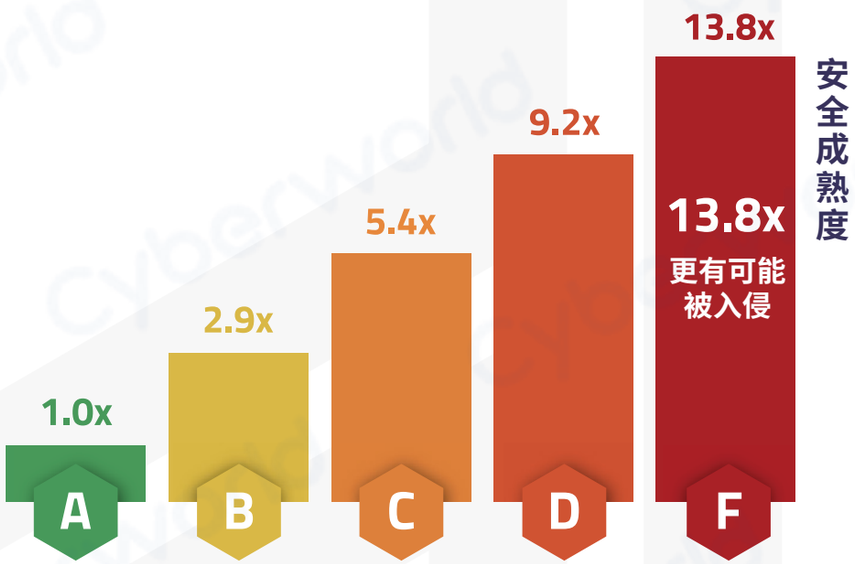
建筑工程业	医疗业	医药业
教育业	酒店业	零售业
能源业	信息业	科技业
娱乐业	服务业	电信业
金融服务业	法律业	运输业
食品业	制造业	
政府	非盈利机构	

与最终用户合作

SecurityScorecard与其用户保持合作关系，以提高对网络风险的认识并报告准确的调查结果。

SecurityScorecard平台为用户提供了 Score Planner 工具，使他们能够以交互方式制定补救计划来提高评分。该工具提出了一种获得更高评分的途径，用户可以根据自己的偏好进行定制。

此外，用户还可以通过在线提交反驳以及适当的证据，对其评分卡上的调查结果提出异议，例如由于补偿控制或归因错误。SecurityScorecard会审查每份提交的反驳和相关支持证据，并在必要时更正和更新评分卡。提交的反驳通常在48小时内被接受或拒绝。如果被接受，评分卡将在48至72小时内更新。



确认

SecurityScorecard的评分算法已成功通过严格的内部验证和确认测试。

验证测试是一个工程过程，用于确定算法的输出是否符合输入。算法要经过一系列统计检验，包括极端情况，以验证其准确性和稳定性。

确认测试确定评分算法是否满足其作为网络安全风险评估工具的预期用途。即，评分越低，发生不良事件的可能性就越大。

在信用评级领域，较低的评级与较高的违规概率相关。对于网络安全评级，较低的评级（较低的评分）应当与较高的数据泄露可能性相关。

SecurityScorecard根据可用的泄露数据，分析了评分与泄露可能性之间的相关性。统计效能受到公开泄露数据量的限制。并非所有机构和企业都有监管义务披露数据泄露事件，因此多达 60% 至 89% 的泄露事件未被报告，这进一步加剧了挑战。

确认测试表明，与评级为 A 的企业相比，评级为 F 的企业发生数据泄露的可能性高出 13.8 倍。

局限性

虽然SecurityScorecard的网络风险评级可以提供有关不同企业的安全态势及其随时间变化的趋势的重要见解，但也存在一些固有的局限性：

- SecurityScorecard 采用“由外而内”的方法，可以非侵入性地、大规模地对企业的网络安全态势进行外部评估。然而，通常无法检测到企业网络内部是否存在补偿控制。在这种情况下，SecurityScorecard可能会报告过低的评分。但是，用户可以通过提交反驳和支持证据来更正自己的评分，以反映补偿控制的存在。提交的反驳通常在48小时内被接受或拒绝。如果被接受，评分卡将在48至72小时内更新。
- 互联网的动态特性也带来了局限性。动态IP可以每天甚至每小时重新分配一次。通信端口可以在不同时间打开和关闭。域名和IP所有权的变化可以随时发生，但需要时间才能在互联网上传播。互联网的动态特性对任何试图描述其当前状态的过程的准确性施加了根本限制。这种努力的结果必然是概率性的，而不是确定性的。对于SecurityScorecard而言，这意味着虽然评分和归因基本正确，但它们总是会以误报和漏报的形式出现一些错误。
  SecurityScorecard开发了一套由机器学习驱动的算法来最大限度地减少这些错误，并不断增强系统架构，以改进更新频率，使归因和评分尽可能保持最新。

常见疑问解答

Q: 评分多久更新一次?

A: 每天更新。

Q: 评分算法多久更改一次?

A: 每三到四年。

Q: 为什么评分会有波动?

A: 评分会在定期的更新频率(每月一次)中略有波动。这使SecurityScorecard能够在动态威胁环境中保持公平的网络安全风险评级,并根据需要引入反映新风险指标的新问题类型,以便让用户及其生态系统更好地了解情况。除了评分更新,对一个企业的评分是一个纯粹确定性的过程。它取决于数字足迹和发现的安全问题数量。如果这些没有改变,那么评分也不会改变。

Q: SecurityScorecard是否会根据企业规模对评分进行标准化?

A: 通常,大型企业比小型企业面临更大的攻击面。SecurityScorecard使用一种原则性的规模标准化方案,为任何规模的企业提供公平的评分。

Q: 评分重新校准的频率如何?我如何知道它们是否会影响我的评分?

A: 每季度进行一次重新校准。如果即将进行的重新校准会影响您的评分,在重新校准日期前四周,您将在平台上看到一个横幅,显示评分变化的影响,并附上一个SecurityScorecard的知识库文章的链接,以获取更多详细信息。

Q: 我在我的数字足迹中看到一个不属于我的IP,我如何相信你们的归因?

A: SecurityScorecard使用大规模自动化流程执行 IP 归因,使用公共 RIR、DNS 和 SSL 数据以及第三方数据源。由于互联网的动态特性,IP可以按天甚至按小时重新分配给不同的企业,IP归因具有根本的概率性特征,无法做到完全无误。一个独立的渗透测试专家团队对SecurityScorecard评分卡的随机样本进行了审核,以客观地确定 SecurityScorecard IP 和域名归因的准确性。他们发现归因过程的准确率为95%,IP地址的正向归因准确率为94%,DNS记录的准确率为100%。

Q: 因素评分不用于计算总分吗?

A: 因素评分表示基于与这些因素相关的问题类型的每个因素的健康状况。总分将根据问题类型的权重计算,因为因素本身没有任何权重。

Q: 因素评分是如何计算的?

A: 因素评分是根据这些因素中的问题类型计算得出的。每种问题类型都有一个权重,该权重取决于其严重程度,会影响因素评分。

Q: 每个因素的权重是多少?如何确定因素权重?

A: 新的评分算法不再有因素权重,总分直接反映问题类型。因素将继续有因素评分,但不再有因素权重。

SecurityScorecard

Cyberworld

关于SecurityScorecard

SecurityScorecard 由 Evolution Equity Partners、Silver Lake Waterman、Sequoia Capital、GV、Riverwood Capital等世界级投资者资助。SecurityScorecard是网络安全评级领域的全球领导者，持续对超过1,200万家企业进行评级。SecurityScorecard 由安全风险专家 Aleksandr Yampolskiy 博士和 Sam Kassoumeh 于 2013 年创立，其专利评级技术被30,000多家企业用于风险管理、第三方风险管理、董事会报告、尽职调查、网络保险承保和监管监督。SecurityScorecard是第一家提供数字取证和事件响应服务的网络安全评级公司，为其全球用户和合作伙伴提供全方位的安全预防和响应方法。SecurityScorecard通过改进向董事会、员工和供应商传达网络安全风险的方式，不断让世界变得更加安全。每个企业都有权获得其可信且透明的即时SecurityScorecard评分。

Cyberworld
广州科明大同科技有限公司

中国区
总代理

官方网站 www.cyberworld.com.cn
业务电邮 info@cyberworldchina.com
服务专线 400-9988-792