DOI:10.3969/j.issn.1009-6868.2016.02.010
网络出版地址:http://www.cnki.net/kcms/detail/34.1228.TN.20160219.1622.002.html

# 大数据安全必须面对的攻击假设矩阵

## Matrix of Attack Hypothesis Faced in Big Data Security

*潘柱廷/PAN Zhuting*
(启明星辰公司,北京 100193)
(Venustech Group Inc., Beijing 100193, China)

中图分类号:TP393 文献标志码:A 文章编号:1009-6868 (2016) 02-0044-005

**摘要:** 认为大数据安全研究需要从大数据攻击研究出发。大数据攻击不仅仅需要考虑针对大数据系统的攻击,更要综合考虑针对系统、过程、数据、语义等多层次的攻击,还要综合看待攻击面和背后的攻击目标。为了更好地理解大数据攻击,提出了意识信息物理系统(MCPs)这样的多层次复杂系统的认识模型,并根据MCPs的多层次,建立起[攻击面x攻击目标]的攻击假设矩阵。对于攻击假设矩阵中每个格子的研究,可以帮助人们构建更有效的保障体系。

**关键词:** 大数据安全;攻击假设矩阵;攻击面;攻击目标;MCPs

**Abstract:** Big data attacks are the foundation of big data security. For data attacks, we need to consider attacks for the systems with large data, and for the system, process data and semantic level of attack, and also need a comprehensive look at the target of the attack surface. In order to better understand big data attacks, a new model of the multi-level complex system—Mentality-Cyber-Physical system/space (MCPs) is proposed. Based on this multi-level of MCPs, a attack hypothesis matrix by [attack surface, attack target] is built up. Research on every grid of the matrix will lead to more effective assurance solutions.

**Key words:** big data security; matrix of attack hypothesis; attack surface; attack target; MCPs

## 1 大数据安全的范畴

大数据作为一个新的技术模式和学科分支,已经开始对网络信息安全产生深刻的影响,这种影响既要循着安全本身的固有规律,也会带着数据自身以前不被重视的新特性。

### 1.1 安全的本质性结构

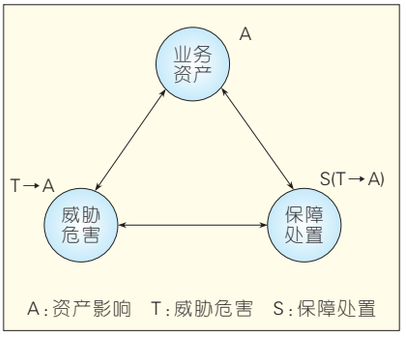
在IT领域的各个分支中,网络信息安全区别于其他分支的根本不同,就是安全永远是一个三要素互相交织、博弈的课题。这3个要素为:业务和资产、威胁和危害、保障和处置,如图1所示。

安全的独特性在于:有难以控制、难以意料的"威胁和危害"一方,自然就有了特有的"保障和处置"这一方,两者和业务资产一起形成了一个三方博弈关系。

所有的安全问题,都要就这3方面分别阐述清楚才谈得到思考的完备性,而大数据安全这个话题也不例外。

### 1.2 大数据安全的方向

大数据安全的如下3个方向,是大数据方法和技术作用于安全三要素所演绎出来的方向。




▲图1 安全的3个要素

(1)大数据作用于业务和资产,即大数据的主流应用。这必然会面临新的针对大数据的攻击和威胁,进而对大数据的保护要对抗这种针对大数据的攻击。

(2)大数据作用于威胁和危害,即大数据攻击和副作用。如果是主动和故意的举措,那就是大数据攻击;如果是被动的,就是大数据产生的副作用,比如大数据技术对于公民隐私保护的破坏。

(3)大数据作用于保障和处置,即安全大数据应用。就是在对抗各类安全威胁的时候,运用大数据技术进行分析和检测,特别是无特征检测、异常检测、态势分析等方面。

文章论述的重点是大数据安全的第1个方向。研究对大数据的保护必须先研究针对大数据的攻击,如果没有真正研究、设计、实现并测度大数据的攻击,那么之前所设计的所

收稿时间:2016-01-10
网络出版时间:2016-02-19

谓大数据防护就都是臆想，只有真实的攻击才能够验证保护和防护的有效性。

## 2 数据本质和特质

研究针对大数据的攻击，我们必须搞清楚针对大数据的攻击的对象——大数据对象。

### 2.1 大数据的7V特性

在描述大数据问题时，我们常说其有7个V的特性[1]，具体如下：

1V（Volume），即海量的数据规模。这体现了大数据问题在数据量上的海量。

2V（Velocity），即快速数据流转和动态数据体系。这代表了时间轴上的大数据，除了对于分析快速及时的要求之外，还体现海量数据可能来自于时间轴的长度延展（存储）和颗粒度的细化（频度）；时间的相关性也是数据间相关性的一大类，比如视频和音频数据就是“顺序时间”的典型结构。

3V（Vast），即数据来自广大无边的空间。每个数据都来自于一个空间的位置，可能是物理空间（现实世界），也可能是网络空间，空间的相关性也是数据间相关性的一大类，也是一大类典型结构。

4V（Variety），即多样的数据类型。大数据，比所谓的“量大”更重要的一个特性就是“高维”。特别是当数据样本的数量难以满足对于高维问题求解的基本要求时，大数据更倾向于回避精确解的求解，而满足于有价值的近似解。这种不追求精确解的特性，让大数据及其系统具有了一定的鲁棒性基础，增加了攻击难度。

5V（Veracity），即数据的真实和准确性更难判断。数据有好坏问题，而这个好坏问题在大数据中会更加极端地被放大，更泛地表达这个话题就是数据的“质”，即数据质量[2]的相关问题。

6V（Value），即大数据的低价值密度。对于大数据的攻击，背后必然要针对其价值进行。

7V（Visualization），即大数据可视化的重要性。大数据的价值需要展现，如果能够破坏和斩断价值链，也是重要的攻击成果。

在这7个V中：第1个V，表达的是大数据外在表现的“大”量；第2～4个V是从时间、空间和多样性这3个方面说明大数据的“大”；第5～7个V阐述的是大数据的价值流转，即从数据本身的客观质量，到有立场的价值认识和价值挖掘，最后到价值的展示和利用。

### 2.2 攻击大数据的常规理解

在传统的网络信息安全领域中（这里指融合大数据特有特征的思考之前），对于攻防的认知主要集中于系统方面：漏洞是系统的漏洞，越权是对于系统访问控制的突破，拒绝服务攻击是对网络系统的拥塞，伪装是对于系统访问者身份的假冒等；安全方法也主要都围绕系统的防护而展开。当然，这个系统是包括了节点式的系统（如主机操作系统）、结构化的网络系统。

在探讨攻击大数据的时候，我们首先想到的就是如何攻击大数据系统，而由于大数据目前的主要应用模式就是分析和决策支持，其系统的对外暴露面非常少，因此至今还没有关于重要的大数据系统遭遇渗透性攻击的报道。能够见诸报道的大数据系统出现的问题和故障，常常是由于电力故障等物理性故障导致的可用性事故，而这些所谓的问题并没有体现出大数据的独特性。

对于大数据系统的、具有针对性的攻击假设，需要针对大数据系统的分布式特色发起攻击。对于大数据的特色攻击还没有太多的研究，可能有两个原因：第一，大数据系统还在快速地演化和发展；第二，攻击研究者要搭建一个接近真实的大数据系统，其成本比较高，技术门槛也较高。但是，由于大数据系统的高价值聚集，这样的攻击早晚会到来。

### 2.3 MCPs结构

网络空间已经成为了大家非常熟悉的一个词，它不仅仅指网络相关的IT系统，更被人们理解为一个空间，在这个空间中主要体现了Cyber实体及其“活动”。

这里所说的活动指Cyber过程，主要体现为操作和流。数据实体对应的是数据流，应用系统对应业务流和服务关系，节点系统对应了计算操作和存储承载，网络系统对应了网络流和连接关系，而物理实体则是对前述Cyber实体的承载。Cyber实体就如同生物体的解剖关系，而Cyber过程如同生物体的生理关系。当前流行概念中的云计算、移动互联网等等都是Cyber自身形态的多样化、高能化和效益化。

信息物理系统（CPS）强调了Cyber与物理空间的关系：可将Cyber与物理空间的关系简化为控制与感知的关系。CPS类似的模型将物理世界和网络空间关联起来了，其关联的根本媒介其实是数据。当前流行概念中的物联网、工业控制、智能生活等等都是将Cyber空间与物理世界更加紧密地关联起来。

网络空间安全领域被分为两大领域：一个是从技术上说的网络安全，比如加解密、攻防渗透、系统加固等；另一个是从系统的内容上说的信息安全，比如舆情态势感知、社交网络策动攻击等。这两方面现在是单独研究和治理的，交集不大。

现在，随着大数据的方法和技术日益得到重视，数据也越来越受到人们的重视。大数据又是一个应用驱动、价值驱动的领域。当数据与数据的语义总是密切关联在一起的时候，我们就发现人的意识空间和Cyber空间的关系变得密切起来。多人的共同意识空间就是群体社交意识。

数据将人的意识空间（包括群体

意识）、Cyber 空间、物理世界 3 方面链接在一起，形成了一个整体意识信息物理系统（MCPs），如图 2 所示。

当我们有了 MCPs 这样的整体认识，在考虑安全问题（特别是大数据安全问题）的时候，就要考虑 MCPs 模式下的攻击。

## 3 MCPs 的攻击假设矩阵

### 3.1 攻击面和攻击目标

攻击面是指攻击者的着手之处和着手模式；攻击目标是指攻击者希望被攻击体系中的某个部分或环节出现重大偏差。我们将攻击面和攻击目标分开来定义，是因为两者并非总是同一的。

### 3.2 MCPs 的 3x3 攻击假设矩阵

在系统攻击中，攻击面和攻击目标可能不同。这种攻击面与攻击目标的错位，可能出现在 MCPs 的 3 个方面，由意识空间、网络空间、物理空间（现实世界）的交叉攻击假设，形成如图 3 所示的 3x3 攻击假设矩阵。

### 3.3 MCPs 的 14x14 攻击假设矩阵

要对 MCPs 攻击假设矩阵进行更具体的研究，就需要将 MCPs 分解成更细致的环节。我们可以将 MCPs 简单分解为 14 个方面，其编码如下：

- Mm：动机
- Mv：价值
- Ms：语义
- Cd：数据和数据流
- Cm：元数据和纯数据
- Ca：应用和业务流
- Cc：计算节点
- Cs：存储节点
- Cn：网络和网络流
- Cp：Cyber 物理实体
- Pc：控制器
- Ps：传感器
- PS：空间关系
- PT：时间关系

将 MSPs 的这 14 个方面组成一个矩阵，矩阵不同的行代表不同的攻击面，矩阵不同的列代表不同的攻击目标。如表 1 所示。

表 1 中，蓝色区域就是从传统的系统攻击视角看到的攻击假设，攻击面可能是网络系统、存储节点、计算主机、应用系统，而最终最受影响的攻击目标也在这其中。

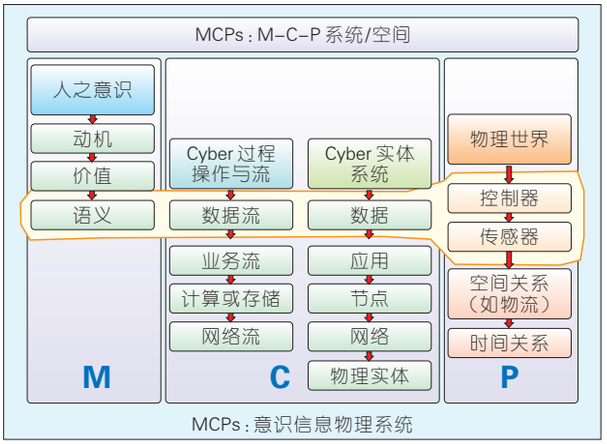



图 2 数据贯穿 MCPs 模型示意

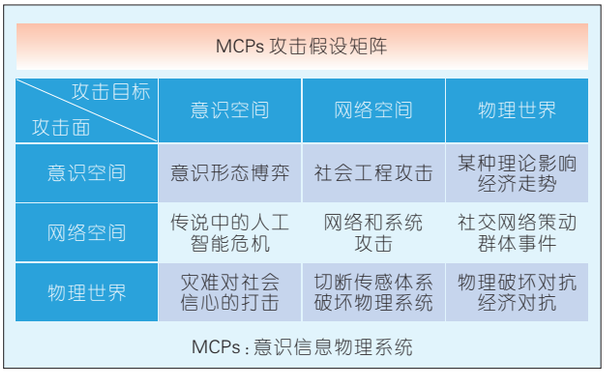



图 3 3x3 攻击假设矩阵及其示例

▼表 1 MCPs 攻击假设矩阵（14x14）

| | Mm | Mv | Ms | Cd | Cm | Ca | Cc | Cs | Cn | Cp | Pc | Ps | PS | PT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mm | | | | | | 4.6 | | | | | | | | |
| Mv | | | | | | | | | | | | | | |
| Ms | | | | | | | | | | | | | | |
| Cd | | | 4.5 | | | | | | | | | | | |
| Cm | | | | | | | | | | | 4.7 | | | |
| Ca | | | | | | | | | | | | | | |
| Cc | | | | | | | 4.1 | | | | | | | |
| Cs | | | 4.5 | 4.4 | | | | | | | | | | |
| 4.5 | | | | | | | | | | | | | | |
| Cn | | | | | 4.7 | | 4.2 | | 4.3 | | 4.7 | | | |
| Cp | | | | | 4.8 | | | | | | | | | |
| Pc | | | | | | | | | | | | | | |
| Ps | | | | | | | | | | | | | | |
| PS | | | | | | | | 4.8 | | | | | | |
| PT | | | | | | | | | | | | | | |

Ca：应用和业务流　Cm：元数据和纯数据　Cs：存储节点　Mv：价值　PS：空间关系
Cc：计算节点　Cn：网络和网络流　Mm：动机　Pc：控制器　PT：时间关系
Cd：数据和数据流　Cp：Cyber 物理实体　Ms：语义　Ps：传感器

注：表格中的数字是文章的章节编号，对应了这个矩阵节点的解释和示例。

数据Cd和元数据Cm，将MCPs三大空间连接起来。表1中的红色部分表示数据作为攻击面和攻击目标会横纵贯穿整个攻击假设矩阵，而且数据会成为MCPs3个空间的桥梁，产生交叉攻击的可能性。

表1反映了大数据和数据视角引入后，给我们带来的更加全面统合的攻击假设视界。

## 4 MCPs攻击假设矩阵的归类分析

MCPs的14x14攻击假设矩阵中的每一个格子，都是一种攻击模式，甚至是一个攻击链的索引。归类后的每个格子，都具有一定的攻击模式共性；格子之间则应当有攻击模式的差异化特点。

做出这样的分类研究，可以让我们把攻击研究得更细致，比如可以将计算节点（Cs）进一步细分为PC节点、移动节点、工控节点等。这样还可提醒我们注意那些原先忽视的空白部分，是否有攻击可能存在。只有对于攻击的全面和细致的研究，才能让我们对于防御和对抗的问题上有更多的把握。

### 4.1 [Cc，Cc]攻击

[Cc，Cc]攻击是最常被关注到的攻击模式，比如，对于操作系统漏洞的挖掘和利用，进而对于系统进行破坏和渗透，其攻击面和受影响目标都是系统。

### 4.2 [Cn，Cc]攻击

与节点攻击不同，[Cn，Cc]对网络的攻击是对结构的攻击。另外，一般把对于网络设备的攻击归类为对于网络的攻击。

分布式拒绝服务攻击（DDoS）是一个典型例子，其通过对于网络结构性的攻击，并通过占领海量节点而构成了一个攻击网络结构，将流量导入给一个目标系统使其瘫痪。这是典型的攻击网络最终危害节点系统。

网络劫持窃听也是一个典型例子，攻击点在网络的路上。通过窃听下来的明文或者密文进行分析，达到渗透相关系统的目的。

从Cn到Cc的影响传递很直接，因为计算节点都自然连接在网络中，所以对网络的攻击会很快传递给计算节点。

### 4.3 [Cn，Cn]攻击

内容分发网络（CDN）是当前一个非常重要的网络服务。如果能够利用CDN服务构建一个CDN指向的环，当向这个环投入足够多的流量时，环就会利用CDN机制在网络中形成一种自激振荡式的流量洪流，可能导致网络风暴的发生[3]。这是典型的攻击网络而危害网络，是一种结构性破坏。

### 4.4 [Cs，Cd]攻击

[Cs，Cd]攻击存储设备，甚至渗透并控制存储设备，自然会对于存储设备上存储的数据产生直接的危害。

### 4.5 [Cs，Ms]:=[Cs，Cd][Cd，Ms]攻击

如果[Cs，Ms]:=[Cs，Cd][Cd，Ms]攻击存储并对存储进行破坏，或者对于存储的攻击和篡改被较快发现，那么这种影响就难于进一步传递到其他攻击假设矩阵格子。

如果对于存储的攻击充分考虑了存储的数据结构，在篡改中保持其基本的数据结构，不让这样的篡改被轻易发现；同时，篡改的数据又能够借助应用系统的分析对于分析结果进行有效影响，那么就能够将这样的攻击传递到语义层，进而影响人的意识空间，影响人的决策。

而如果要在大数据存储环境下达到[Cs，Ms]，就要顺应大数据存储的系统模式和其存储数据的数据结构，做到篡改不易被发现；还要了解大数据存储的数据将如何被分析和应用，让篡改的数据能够污染到大数据分析的结果。

大数据相关的攻击假设，能够让我们反思如何对抗这种攻击。如果将存储的系统模式和数据结构进行一定的随机化（仿效操作系统中的地址随机化思想），那么大量篡改数据就很容易被发现；如果将大数据分析的容错能力（容忍不良质量数据）提高，那么就迫使要污染大数据分析结果必须篡改更多的数据。让“篡改不易被发现”与“大量篡改数据才能产生语义污染”形成矛盾，进而将攻击的效果阻隔在Cyber空间中，不让其有效影响人的意识空间。

### 4.6 [Mm，Ca]攻击

2016年初的一个突发案例[4]：一则谣言，经过微信朋友圈的扩散，震动了大半个互联网金融圈。

2016年1月10日下午，回顾2015年微信数据的“我和微信的故事”在朋友圈突然被刷屏，正当大家玩得非常欢快时，一个哑弹突然向社群中抛来。当晚，有用户在自己的朋友圈中称：该链接“千万不要进，（黑客）马上把支付宝的钱转出去，已经有人被盗”，还称加载该链接时“很慢，已经在盗取资料。”朋友圈截图被疯转，引发用户集体不安。很多人吓得把支付宝的银行卡都解除绑定，支付宝里的余额全部打回银行卡，还一一提醒朋友“如果我这个号向你借钱，千万别理。”

在1月11日的一个报告中，张小龙说起10日晚的事称：“我和微信的故事”的链接没想到被分享出去，这样带来了3个问题。第1个问题：访问太高，基本挂掉了；第2个问题，有人造谣说，打开链接支付宝的钱被偷了，这个时候，链接也确实因访问量太高打不开了；第3个问题，百万级用户开始解绑银行卡了，结果服务器也快挂了，银行卡也解绑不了了。

这是一个典型案例：一个谣言（在人的群体意识空间），影响了人们的操作行动，进而让一个应用系统崩溃（网络空间中）。

对于这类有意的攻击和无意的危害，有些防范措施可能在意识空间，有些防范措施就要在网络空间，甚至需要二者结合。比如，针对这类[Mm，Ca]风暴，就可以考虑建立态势感知监控和相关性研判，当然这就要将舆情监控和系统风暴监控进行相关性联动分析。这在以前是没有的，从这个事件让我们意识到这种联动分析的必要性。

### 4.7 [Cn，Pc]:=[Cn，Cm][Cm，Pc]攻击

光大证券乌龙指事件[5]给我们展示了一种可能性。

2013年8月16日11点05分上证指数出现大幅拉升，大盘一分钟内涨超5%，最高涨幅5.62%，指数最高报2 198.85点，盘中逼近2 200点。11点44分上交所称系统运行正常，下午2点，光大证券公告称策略投资部门自营业务在使用其独立的套利系统时出现问题。有媒体将此次事件称为“光大证券乌龙指事件”。

一个系统网络的故障，可能导致应用系统和大量数据的错误，这些可能是数据Cd或者元数据Cm。如果一些金融衍生品应用系统是通过对数据监测和分析自动进行买卖操作的，就可能因为被监测数据的错误导致错误的买卖决策（控制现实世界的控制器行动）；而如果错误的买卖决策又继续导致被监测数据的错误效果放大，可能就在市场中产生连锁效应，甚至有引发或诱发证券市场的瞬间大波动甚至股灾。

这种危害的可能性，对于社会的危害是极为严峻的。

### 4.8 [Cp，Cm]攻击和[PS，Cs]攻击

在密码破译和密钥分析领域，有一种方法：通过对密码芯片外部的热量分布进行跟踪分析，从而达到破解和猜测密钥的目的。这是典型的[Cp，Cm]攻击，用对系统物理实体的分析来攻击到数据层。

对于系统的运行状态进行分析，我们也可以通过对系统的能量消耗进行分析。这是典型的[PS，Cs]攻击，用物理世界的物理测度PS来分析系统Cs。

上述两个分析（攻击）都需要对物理世界测度并产生相当大量的数据，才能完成对于Cyber内部的分析。换句话说，这个分析过程需要大数据技术和分析方法的支持。

## 5 结束语

MCPs攻击假设矩阵还有很多空白之处需要填补和研判。可以想象：当我们把各个格子的攻击都能够假想并模拟出来，那么对于有效的安全保障和问题防范就会产生不可估量的支撑。

大数据安全绝对不能停留在系统层面，一定要在MCPs的统合视角下研究整个攻击假设矩阵。特别是跨MCP三大空间的攻击，将是非常值得研究的，很多“黑天鹅”式的攻击必然由此而产生。

**参考文献**

[1] 潘柱廷. 安全大数据的7个V——大数据基础问题与信息安全的交叉探究[J].中国信息安全, 2013(9):74-77
[2] MCGILVRAY D. 数据质量工程实践[M]. 刁兴春, 曹建军, 张健美, 译. 北京: 电子工业出版社, 2010
[3] CHEN J J, JIANG J, ZHENG X F, et al. Forwarding-Loop Attacks in Content Delivery Networks [EB/OL]. [2010-12-10]. http://netsec.ccert.edu.cn/duanhx/files/2010/12/cdn_loop-final-camera-ready.pdf
[4] 微信之父张小龙：“微信盗号谣言”引发蝴蝶效应[EB/OL]. [2016-01-11].http://news.ifeng.com/a/20160111/47022386_0.shtml

**作者简介**

**潘柱廷**，教授级高工，启明星辰公司首席战略官，中国计算机学会CCF第十一届常务理事，CCF大数据专家委员会专家，CCF计算机安全专业委员会常务委员等；长期从事网络信息安全技术的研究、开发，以及公司的战略研究策划、技术管理等工作。

# 综合信息

### 全球公共云市场规模2016年将达2 040亿美元

市场研究公司Gartner发布报告称：全球公共云服务市场规模2016年有望达到2 040亿美元，较2015年的1 750亿美元增长16.5%。

据预计，公共云服务市场将继续呈现出高速发展态势，并一直持续至2017年。Gartner公司指出：虽然公共云服务呈现出稳定发展的态势，但是2016年发展速度最快的却是IaaS。2016年，IaaS有望增长38.4%；到2016年年底，IaaS市场规模有望达到224亿美元。

此外，SaaS也有望实现年增长20.3%，达到377亿美元。云管理和安全服务的增长率有望达到24.7%。PaaS也有望表现出非常强劲的发展势头，达到21.1%的增长率。

市场研究公司ZK Research的分析师宙斯·科拉瓦拉则预测：云服务呈现出的这种强劲发展势头有望在未来5～7年内仍然保持下去。“我想我们尚处于云服务时代的起步阶段。”他表示，“我们可能会继续看到越来越多的公司逐渐向云端转移。很多公司因为担心安全问题而对云服务避而远之。但是，随着时间的推移，人们的这种担心最终会淡化乃至消失，越来越多的企业将会增强对云服务的信心，而这必将会继续促进云服务市场的发展。（转载自《中国信息产业网》）