数据安全治理的基本思路

杜跃进

2018-11-07

我们的世界正在进入一个奇怪的分裂状态：一方面人们为大数据时代即将在各个领域发生的革命性进步而激动难眠，一方面人们也在为数据安全和隐私保护问题担心得睡不着觉。围绕大数据的创新和安全，各种政策、法律、标准、产品和学术研究表现出空前的热情。然而眼花缭乱的声音却使人们陷入了混乱，陷入了数据恐慌。如果我们不能尽快找到清晰的思路，不能尽快找到方法实现围绕大数据的发展与安全之间的平衡，我们可能丧失人类历史上迄今为止最大的一次发展机会，或者陷入最大的安全危机。本文要讨论的，就是大数据时代下该如何进行数据治理的基本抓手与核心思路。

一、 若干基本认识

关于数据安全，我们需要首先建立一些基本的认识。

首先，数据本身无罪，有罪的是数据没有被安全地保护或使用。大数据时代，每个角落都在产生数据，而这些数据正是新时代人类的财富：我们不仅依靠这些数据提供更精准贴心的服务，更依赖这些数据实现医疗、健康、教育、安全、环境保护等各方面的革命性进步。可是人们担心个人隐私在这个过程中被窥探，对似乎无所不在的数据采集记录行为无法忍受。但是回想一下，人们曾经在日记里写下自己最私密的事情、人们在自己的手机或计算机中存储自己的私密照片或信件、人们在很多

政府部门的系统里存储自己的各种生物特征信息、人们在医院的系统里存储自己各种病情细节、人们向心理医生倾诉自己的内心等等，这些时候为什么不担心隐私泄露呢？因为这些情况下我们觉得自己的隐私数据是不会被泄露的，虽然事实上也存在风险。所以，其实数据本身是无罪的，人们担心的是数据拥有或者处理者不能保护好数据或者滥用数据。因此，今天我们谈到隐私保护的时候，不应该只关注法律层面的隐私条款和限制信息采集，而应该更加优先关注如何提高所有拥有我们数据的组织的数据安全水平，确保他们手里的数据不会被窃取或者滥用。“用户画像”、“精准营销”不等于就会侵犯隐私，关键看具体实现的方法和管理措施。

其次，谁都不信任的话，用户的安全可能会更加糟糕。越来越多的人似乎倾向于“谁都不信任”，甚至一些研究也在朝着这个方向努力。但是在今天除非你不生活在人类社会中，否则这种思路反而让你陷入到更加不安全的地步。私密数据放在自己的手机上或者服务器上就更安全吗？除非你的手机或者服务器从来不用任何形式和网络发生链接，否则对网络攻击者来说，这些安全防护相对更弱的地方恰恰是更容易的目标。当然，如果你就认为自己能够永远打败全世界所有的网络攻击者那也行。那么不使用任何电子产品不行吗？就算假设未来这样做依然可以生活，答案也是不行，因为用电的数据、看病的数据、出门走路的数据等都会被别人记录在别的地方。所以，如果我们不相信所有提供服务的企业或组织、我们也不能相信第三方机构因为他们的安全能力未见得更好、我们也不能相信自己能够抵抗全世界的网络攻击者，那怎么办？

第三，安全也需要数据。我们担心泄露了自己的数据所以不安全，可是反过来，如果没有数据的话我们也无法得到安全保护。徐玉玉案件因为坏人偷了她的相关数据而产生了危害，章盈颖案件我们则多么希望好人知道她的位置信息从而能帮助到她。追踪老人或者儿童的位置信息可以防止他们走失，野外应急救援也需要位置信息，急救大夫需要病人的隐私健康信息才能救命，通过检测用户是否短时间在不同的城市登录系统是今天几乎所有产品判断用户账户是否被盗的基本手段……

第四，我们要解决的是“大数据时代下的数据安全”，而不是狭义的“大数据安全”。包括徐玉玉事件在内的各种案例，实际上都不是从所谓的“大数据”那里偷取数据的。网络空间不存在物理位置的限制，因此现实中攻击者更容易从各个安全薄弱的服务器或组织那里下手，而不是和防护严密的大型大数据公司对抗。以电商为例，猖獗的黑灰产绝大多数都是瞄准商家、独立软件供应商、物流等各个环节下手窃取订单等信息，然后用于诈骗等活动。大数据时代，数据在开放、复杂、快速变化的业务和产品链条中高频流动，数据成为复杂生态的每个环节中都无法剥离的部分。这是导致数据安全问题变得空前突出的根本原因，因为所有过去的数据安全方法基本上都失灵了。

二、 大数据时代的数据安全包含哪些内容

目前存在的比较普遍的误区是把系统安全等同于数据安全，也就是说，把防止网络入侵带来的数据被窃取，等同于数据安全的工作。实际上这只是数据安全很小的一部分内容。

今天我们说数据安全的时候，应该包括三方面的内容：防窃取、防滥用和防误用。

防窃取比较容易理解，不过全世界多年来的共识是，来自内部的安全威胁总体上占三分之二左右，要远大于来自外部的威胁。根据电子商务生态安全联盟的统计，在电商生态中这个比例还要更加悬殊。因此，即便是系统安全，也不能仅仅把防止外部攻击导致的数据窃取作为全部工作，防止来自内部的入侵和数据窃取反而更加重要。

防滥用指的是防止数据被不正当使用，例如拥有数据的组织内部员工在无工作场景的情况下访问用户个人敏感数据。现实中，用户的身份证信息、医疗档案、购物记录、财产情况等信息，都会存在各种组织的系统中。当用户需要这些组织提供服务或者帮助的时候，这些组织的相应员工才可以根据用户的授权来访问这些数据。而如果这些组织中的员工未经用户请求私自访问这些数据，则属于一种滥用行为。从已经破获的并且披露的众多电信诈骗案件中可以看出，大量内部人员滥用职权倒卖用户信息，这些都属于数据滥用的场景。目前大部分组织对这部分工作的意识淡薄、能力不够。在技术上是能够实现这类行为的监控的，配以制度的保障，可以有效遏制这类滥用行为。有些业务场景更加复杂一些，例如包裹邮寄单上显示的收发件人的详细信息，在整个包裹流转过程中都面临泄露风险（现实中这些信息都是网络黑灰产收购的对象）。但即便这类场景，也有“隐私面单”等相应的技术出现。防滥用也包括一个组织对外进行数据披露、数据共享、数据转移等过程中的安全审核，这些审核工作确保这些行为合法，并且不会导致用户或者组织自身的利

益受损。脸书事件最早的问题就是出在这个环节。

防误用指的是防止数据在加工过程中出现过失性泄露。人类正在进入定制化生产的时代，这个时代的基础之一是基于大数据的加工计算。大数据加工计算的过程中如何做到不侵犯用户个人隐私，就是典型的防误用问题。显然这个问题已经成为今天的典型突出问题了。实际上人们今天谈之色变的“用户画像”、“精准营销”等，早已经在普遍使用了，而且这些都是未来数字经济、智慧城市和治理等工作必不可少的技术。只是到具体的实现层面，有没有采用合适的制度和技术手段确保这些过程中不会泄露特定人的隐私，是今天每个组织需要回答的数据安全问题。在技术上这也不是无法做到的，有很多比较成熟的方法可以做到让用户感觉有个贴身秘书在给自己服务，但是实际上没有人能够在数据加工的全过程中窥探到特定用户的隐私。可是在意识上，恐怕现在绝大多数组织在这方面还需要提高。

三、 数据安全治理三原则

搞清楚数据安全要解决哪些问题、大数据时代下解决这些问题所面临的主要挑战，就可以梳理数据安全治理的核心思路了。简单说，数据安全治理可以遵循“以数据为中心、以组织为单位、以能力成熟度为基本抓手”的原则。

1、以数据为中心

以数据为中心，是数据安全工作的核心技术思想。人们比较习惯的是以系统为中心的思想，即围绕着一个数据库、一个产品、一个网站、

一个服务器等评价其安全性。这种思路主要适用于保护一个特定系统的正常工作状态。但是在今天，数据在多个系统、产品、业务环节中频繁快速流转，这种以系统为中心的思想已经不能满足数据安全的需求了。以数据为中心的安全，是将数据的防窃取防滥用防误用作为主线，在数据的生命周期内各不同环节所涉及的信息系统、运行环境、业务场景和操作人员等作为围绕数据安全保护的支撑。这时候，某个系统被入侵，并不等于数据安全的目标就遭到最终的破坏，反之某个单一环节的安全能力再强，也不代表整体数据安全保护的能力就够好。

在数据生命周期的不同阶段，数据面临的安全威胁、可以采用的安全手段有可能很不一样。例如，在数据采集阶段，可能存在采集数据被攻击者直接窃取，或者个人生物特征数据不必要的存储面临泄露危险等；在数据存储阶段，可能存在存储系统被入侵进而导致数据被窃取，或者授权用户无应用场景支持访问用户敏感数据，或者存储设备丢失导致数据泄露等；在数据处理阶段，可能存在算法不当导致用户个人信息泄露等。把不同阶段从不同角度面临的风险放到一起进行综合考虑，建立强调整体而不是某个环节安全能力，是以数据为中心的安全的核心思想。

2、以组织为单位