调查观察 周刊

本刊主编：易艳刚
值班责编：刘晶瑶、雷琨
实习生：王晓东

第45期

微信公号：xhmrdxwx 电话：(010)63076340

别让玻璃幕墙成“不定时炸弹”

数据库“裸奔”，个人信息屡被窃后在“暗网”挂售

不法分子利用一些机构松松垮垮的保护意识，大肆窃取数据牟利

“暗网”可将传输的数据进行加密，并且在数据传输过程中，利用其他“暗网”节点进行多次随机转发，从而实现了信息发布者身份信息的隐藏或保密。因此，最终的信息接收者虽然能收到信息，却很难找到信息的发送源头

一些机构认为自己的主业不在线上，遭黑客入侵的可能性不大，但酒店、航空、教育培训等传统行业的数据信息恰恰是地下黑产的最爱

本报记者颜之宏、关桂峰

漫画：曹一

在“黑色产业圈”，**拖库**意指将机构数据库中的重要数据窃走。由于拖库与“脱裤”发音接近，且重要数据中包含了大量用户隐私信息，因此这个词还暗喻被窃取隐私信息的用户被扒得“一丝不挂”。

今年以来，多家机构的用户数据库发生拖库事件，包括网购商品信息、学籍信息、个人从业经历甚至开房记录等**高度敏感信息均在“暗网”上挂售**。不少人提出质疑：我们究竟还有什么隐私没有被泄露？把隐私交给互联网企业究竟安全吗？

“大门敞开”的数据库与黑产的狂欢盛宴

“出售华住集团旗下所有酒店数据(汉庭、美爵、禧玥、漫心、诺富特、美居、CitiGo、桔子、全季、星程、宜必思尚品、宜必思、怡莱、海友)，附件当中为测试数据，各提供10000条数据供大佬测试……”8月28日，一张经由“暗网”流出的截图在社交媒体上疯转，有地下黑产从业者声称掌握了华住集团旗下酒店近5亿条数据信息，并以打包价8比特币或520门罗币(当时折合人民币约37万元)公开出售。

一石激起千层浪，不少人开始在朋友圈感叹“在互联网时代毫无隐私可言”，也有人质疑在“暗网”出售的数据并非真实信息。然而，第三方网络安全团队对“暗网”公布的3万条数据样本进行技术鉴定后认定“样本数据准确”。

更让人惶恐的是，在“暗网”挂售数据的地下黑产还表示，“以上数据获取时间为2018年8月14日，如果权限不丢失，后续数据还可以免费发给已购买上述数据的买家”。也就是说，地下黑产对数据库的入侵行为并非“一次性”行为，而是获取了访问数据库的权限，如果有关方面不采取进一步补救措施，发帖者甚至可以做到在数据库中“来去自如”。

“一些机构认为自己的主业不在线上，也非互联网行业的主要参与者，因此认为自己遭黑客入侵的可能性不大，从而降低了对网络安全防护的要求，甚至不配齐相应IT部门的人员，从而成为网络攻击的受害者。”资深网络安全专家Mystery向新华每日电讯记者表示，酒店、航空、教育培训等传统行业的数据信息恰恰是地下黑产的最爱，“由于线下强制实名制的要求，这些领域的数据真实度很高，也可以更精准地绘制出用户画像，从而给不法分子进一步侵害用户提供机会。”

有业内人士认为，仅就目前在“暗网”挂售的数据来看，黑色产业链从业者自己就可以很精准地做出用户画像——你从什么学校毕业、学的什么专业、曾在哪些机构工作过、喜欢去哪里玩、喜欢买什么、爱看什么类型的电影，甚至和谁住过一间房，都能被大数据精准地“画”出来。

获得上述这些数据并非难事，甚至都不用花费太多的时间和金钱成本。南方都市报个人信息保护研究中心发布的《2017个人信息保护年度报告》显示，**黑市上可以轻易买到搜索对象的个人信息**，其中包含近半年来的通话记录、出行信息、账单消费以及人脉关系等，甚至还有针对搜索对象详细的量化评分。而获取上述详细信息的金钱成本，仅仅需要3.8元。

“与其说是地下黑产猖獗，倒不如说这是黑产们的一场狂欢。”Mystery认为，在互联网时代，用户数据变得越来越“值钱”，**如果有关机构再不把好数据库的安全关，未来“暗网”上将有越来越多的用户隐私被“明码标价”地售卖**。

并不神秘的“暗网”世界与松松垮垮的保护意识

不少人对前文反复提及的“暗网”并没有特别具象化的概念，只是单纯地认为“暗网”就是一个“地下黑市”。

据360行业安全研究中心主任裴智勇介绍，“暗网”的典型代表就是“洋葱网络”，它由美国军方研发并申请专利。简单地说，“暗网”就是将传输的数据进行加密，并且在数据传输过程中，利用其他“暗网”节点进行多次随机转发，从而实现了信息发布者身份信息的隐藏或保密。因此，最终的信息接收者虽然能收到信息，却很难找到信息的发送源头。

不过，裴智勇说：“基于现代网络技术和大数据技术，‘暗网’的追踪溯源已从‘理论不可行’变为‘实际可行’。**一些在‘暗网’上倒卖用户数据的卖家已被警方追踪并抓获**，这说明在‘暗网’上犯罪也不是绝对安全的。”

需要说明的是，**“暗网”并非独立存在的网络**，它也是由运行在普通互联网上的软件或设备组成。只是这些软件或设备遵守“暗网”的通信协议，可以各自独立工作并互联互通，不需要任何管理者就能组成一个“暗网”网络。

裴智勇认为：“从单纯的技术角度看，很难说‘暗网’是好是坏，但从后来的实践来看，‘暗网’并没有像其原始设计者们所想的那样被用于保护公民言论自由，而是被犯罪分子大量用于个人信息、人体器官、武器军火和毒品等非法交易。”

但是，裴智勇也指出：“把‘暗网’等同于黑暗的网络也是片面的，因为实际上，通过普通互联网进行的各类非法交易，规模远远大于‘暗网’交易。”

事实上，在哪里贩卖用户数据并不重要，地下黑产是如何获得用户信息更值得普通用户深思。

“我们过去在办案中发现，一些群众尤其老年人防范意识薄弱，看到大街上在摆摊的有小礼品送就会去填写手机号、身份证号等个人信息，还会在对方的指导下扫描二维码，这都很容易造成个人信息泄露。”厦门市反诈骗中心民警洪恒亮告诉新华每日电讯记者，在一些电信网络诈骗案件中，不法分子冒充领导、亲友或冒充“公检法”查案，很容易就报出了当事人的身份信息，实际上这些信息都是大家在日常生活中无意泄露出去的。

还有一种套取用户信息的方式，则更符合年轻人的生活习惯，那就是在朋友圈中参与“钓鱼活动”。洪恒亮表示，一些营销号会发布诸如“免费酒店试睡”“转发抽锦鲤”等“钓鱼活动”，不少年轻用户缺乏判断力，容易跟风转发从而参与进去，除了填写身份信息外，还“欢天喜地”地转发给朋友圈的其他好友，进一步扩大受害范围。

“参与这类钓鱼活动，轻则成为营销号的‘僵尸粉’，重则接到精准的电信网络诈骗‘全家桶’。**一些不法分子通过对用户提供的身份信息的解读重构，甚至还可以盗取其社交账号**。”洪恒亮说，这些被搜集走的个人信息还可能为犯罪分子利用短信嗅探设备进一步盗取金融账户余额提供便利。

隐身“暗网”的幕后黑手与“上下失守”的防范措施

相较于开房记录这类私密性极高的个人信息对当事人的“杀伤力”，被泄露出去的学生学籍信息则对家长的“钱袋子”更有威胁。

7月30日，一条题为《浙江省1000万学籍数据出售》的帖子在“暗网”某中文论坛中引发关注。发帖者称，其所出售的数据包含学生姓名、身份证号、学籍号、学校名称、学校序号、班级号、户籍信息、监护人姓名、监护人手机号、居住地址和学生照片信息，作价0.02比特币(当时折合人民币约1000元)。

为了取信买家，发帖者还放出一张20多条由上述信息构成的“样本截图”，生源地分别为浙江的杭州、嘉兴、温州等地。新华每日电讯记者随机抽选了3条信息进行电话核实，证实信息准确无误。

或许是信息过于准确翔实，其中一位家长在通话中“执着”地认定记者就是其孩子的班主任，将要对其进行家访。在反复说明后对方仍存疑的情况下，记者不得已要求对方以信息泄露为由报警。

幸运的是，公安机关9月发布通报称，金华市公安局江南分局已于8月10日成功抓获了**非法侵入浙江省学籍管理系统的王某禾等犯罪嫌疑人3名**，查获公民个人信息1100余万条。

今年6月以来，“暗网”上针对我国各政企机构的数据倒卖事件呈多发态势，且多发于敏感事件节点，有专家认为，此类事件背后的问题值得关注。

在诸多机构被拖库事件中，泄露数据准确率较高，所涉数据库种类繁多，其中所展现的黑客挖掘能力较强。北京邮电大学网络空间安全学院副教授芦效峰认为，“暗网”集中出现针对我国政企机构的用户数据倒卖事件或有外部势力支持。芦效峰表示，一些西方国家过去曾在国际场合中多次对我国网络安全环境“指指点点”，在当前复杂多变的国际局势下，有必要提防一些外部势力有组织有计划地实施网络攻击行为。

部分机构数据库维护权责不清，责任主体思想懈怠情况突出。裴智勇表示，**对攻击预警不在乎，对管理规定不遵守，对应急方案不执行，对风险提示不满意**，部分机构安全团队在思想上麻痹懈怠，甚至在数据库泄露后仍不执行应急预案，在筑牢防范网络攻击意识关上“上下失守”。

行政处罚措施落实不及时，在行业中易形成消极氛围。新华每日电讯记者在查阅公开资料后发现，自今年6月A站(AcFun)用户数据库被拖库以来，尚未发现有行政主管部门对有关机构进行行政处罚的通报。专家认为，如行政处罚措施不及时跟进，类似事件或将反复发生。

急需明确的责任主体与亟待建立的应急机制

专家建议，针对暗网上日益频繁、规模愈发庞大的用户数据倒卖情况，有关部门应建立响应机制，同时厘清权责关系，让广大群众在享受互联网技术的同时更多收获安全感。

“无论是拥有法律强制力的网络安全法，还是对行业起约束作用的《信息安全技术个人信息安全规范》，我们针对用户数据保护的法律法规是有的，但是这却没有阻挡住猖獗的倒卖数据的行为，这背后的原因值得有关方面深思。”中国信息安全研究院副院长左晓栋认为，我国在公民个人信息保护上的立法已相对完善，现在需要看到相应部门来落实法律执行。

左晓栋表示，在当前环境下，**有多个部门对个人信息保护负有责任，“九龙治水”情况较为突出**。他建议，可设立专门机构来应对个人信息泄露问题，对此类专门机构赋予相应的司法和行政权力，对数据保护不力者形成震慑，促使行业内形成“用户数据就是机构生命”的良性氛围。

一些第三方网络安全机构在问卷调查中发现，**不少机构并没有建立相应的网络安全应急机制**，在极端环境下缺乏应对措施。裴智勇建议，有关部门应督促相关涉事机构加快建立针对数据库泄露的网络安全应急机制，加大对一些网络安全防护能力不足的机构的指导，倒逼其“防有所指、防有所用”。

此外，部分网络安全专家向记者表示，国内**一些机构为压缩成本，未按要求配齐网络安全团队，导致数据库长期处在“放养”甚至“裸奔”状态**，危险系数较高。专家建议，有关部门要有针对性地对传统行业的数据库进行排查摸底，对不符合信息安全等级保护规范的机构开出负面清单，并责令其限期整改，切实保护用户数据安全。

搬点评·抢车票·“爬”隐私：“爬虫”滥用成害虫

本报记者鲁畅、王阳、颜之宏

近日，有自媒体披露在线旅游网站马蜂窝旅游网涉及产品点评抄袭甚至作假行为。马蜂窝回应表示，将针对审查漏洞采取积极改进措施，但对于歪曲事实的言论和已被查证的有组织攻击行为将采取法律手段维护自身权益。民事诉讼事件有待司法机关调查，但业内人士表示，这一事件折射出技术伦理和法律问题。

新华每日电讯记者调查发现，近5年，互联网行业中用户生产内容平台(UGC)中数据造假情况长期存在，尤其是**网络“爬虫”技术的非法操作**不仅侵犯相关平台知识产权和消费者合法权益，还可导致平台上的用户敏感信息泄露。

捅“马蜂窝”之后：UGC数据造假受关注

日前，微信公号“小声比比”发布题为《估值175亿的旅游独角兽，是一座僵尸和水军构成的鬼城？》的文章，援用乎睿数据团队所供数据称，作为马蜂窝核心资产之一的2100万“真实点评”中，**有1800万条是通过机器人从携程等竞争对手那里抄袭过来的**。其中，超过7000个抄袭账号，合计抄袭572万条餐饮点评，1221万条酒店点评，占总点评数的85%。

随着事件日益发酵，UGC平台数据造假这一行业问题引发关注。业内人士认为，从其他平台抓数据的目的，就是为了制造流量很大的假象，既给用户看、商家看，更要给投资人看，获取不同轮次的投资以便上市。而从其他网站抓取页面商家内容和用户点评数据非常简单，**使用“爬虫”技术和人工编辑就能做到**。

一位App研发者李滨介绍，爬虫最早应用在搜索引擎领域，爬取网站页面提供给其他用户进行快速搜索和访问，当前爬虫技术已是“大数据”概念的重要组成部分，爬取对象也从一些种子扩充至整个网络数据。为此，行业还达成了Robots协议，形成互联网行业就抓取数据普遍遵守的规则。

然而，近年来，一些公司开始利用“爬虫”技术从其他平台恶意抓取数据。例如今年7月，生活分享平台“小红书”官微发布声明指责大众点评大量抄袭小红书用户的内容，随后，大众点评道歉；今年2月，视频弹幕网站哔哩哔哩大量用户的视频、昵称、头像及用户评论，出现在某新成立的视频网站上；而航空公司的官网上的机票、订座等信息，长期被代理公司将机票信息爬取、占座，然后在其他网站上加价销售。

“爬虫”被滥用，数据造假已成网络“灰产”

网宿科技发布的《2018上半年中国互联网安全报告》显示，今年上半年，Web应用攻击总数环比增长了97.82%，恶意“爬虫”攻击数量环比增长了55.79%。另有数据显示，交通出行类恶意“爬虫”流量占比居首位，其次是电商、社交、点评、运营商、公共行政等，网络爬虫的非法使用给互联网竞争环境带来诸多负面影响。

“目前，市面上大的互联网公司都会推出自己的刷票软件，目的是为了分享12306网站的‘流量红利’。”中国铁科院电子所相关负责人说，“**这些刷票软件用‘爬虫’等技术刷新12306网站页面**，截取官网车次、票量等数据制作成自己的网站页面，再使用程序进行抢票，收取不合法的差价。其中，大约一半的12306订票网站流量来自‘爬虫’技术支撑的刷票软件，不仅给网站服务器造成巨大压力，也**扰乱了正常的订票秩序**，由此带来的购票难是铁路部门一直头疼的问题。”

山东日中律师事务所律师陈冠汶说，公司未经许可或授权的情况下利用“爬虫”技术获得可能带来商业利益的信息可以被判定构成不正当竞争。新华每日电讯记者梳理相关案件发现，法院往往认为，技术作为一种工具手段在价值上具有中立性，但这并不意味着技术本身可以作为豁免当事人法律责任的依据。

—— 2017年，广东省深圳市中级人民法院审理的一起案件中，武汉元光科技有限公司为提高其开发的智能公交“车来了”App在中国市场的用户量及信息查询的准确度，未经深圳市谷米科技有限公司许可，指使公司员工利用网络“爬虫”软件获取谷米公司服务器中的实时数据，谋取该软件在实时公交信息查询软件中的竞争优势，违反了诚实信用原则和公认的商业道德，构成不正当竞争。

—— 2016年，上海知识产权法院二审民事判决书认为，百度公司大量使用大众点评网的点评信息的行为，通过百度地图和百度知道与大众点评网争夺网络用户，会导致大众点评网的流量减少，同时，又推介自己的团购等业务，攫取了大众点评网的部分交易机会。百度公司的行为损害了汉涛公司(大众点评网所属公司)的利益，且其行为违反公认的商业道德，构成不正当竞争。

搜狐视频高级主管闵博认为，滥用网络“爬虫”还有可能对网络安全造成影响，引发网站服务器宕机。“一些‘爬虫’工具的使用者在采取‘全站爬取’的模式时，相当于模拟了大量用户在短时间内对源站服务器发起访问请求，一旦访问量在瞬间达到承载极值，就会引起服务器宕机，从而威胁网络空间的安全。”

还有业内人士指出，被网络“爬虫”抓取的信息不仅可以用于同类型平台制作，还可能被转售或者可能用于钓鱼网站制作等其他违法行为，不但会给平台带来重大损失，更可能导致平台上的用户敏感信息泄露，进而使用户遭遇各类网络和电信诈骗。

记者在社交网站和购物网站检索时发现，一些卖家堂而皇之地售卖“爬虫”自动评价软件或进行个人隐私信息爬取的接派单任务。在QQ群搜索中输入关键词“网络爬虫”，也会出现多个涉及外包网络“爬虫”技术的群组。业内人士透露，这些群组中进行的网络“爬虫”任务大部分都属于未获授权而进行的违规爬取操作，由于该项技术具有一定的隐秘性，在爬取普通用户隐私数据时较难被识别。

技术加法律遏制造假，勿让“爬虫”成害虫

“一个技术如何使用，责任主体都应扪心自问，这是否侵犯个人隐私，是否破坏言论自由，是否损害公共利益，是否损害其他数据财产拥有者的财产所有权，是否涉及不正当竞争。许多大的互联网公司，会同时有‘爬虫’部门和‘反爬虫’部门。**‘爬虫’技术在互联网行业早已被广泛使用，但‘爬虫’绝不能成为害虫**。”中国人民大学法学院教授刘俊海表示。

记者了解到，**目前的“反爬”技术有两种**：一种是限制同一IP、同一电脑在一定时间内访问网站的次数，另一种是设置复杂的验证码机制，让“爬虫”不好识别。但对一些网站来说，封IP的做法可能误伤真实用户，而设置一个非常复杂的验证码，又可能损失用户。因此，除了加大技术防范力度之外，要运用好法律手段，克服取证难的盲点，明确使用红线。

业内人士表示，虽然网络安全法对非法获取个人信息等相关行为进行了规定，但对于爬取公开信息行为并未予以规定。相关部门应进一步查漏补缺，尽快缩小新技术应用的法律模糊地带。

据中国传媒大学教授、大数据挖掘与社会计算实验室主任沈浩介绍，欧洲已出台GDPR《通用数据保护条例》，想要采集欧盟境内企业和个人的信息，即使你不在欧盟边界内，采集活动也要受到相应的管制和控制。但目前国内还没有全面的规定，要从根源上解决这类问题，还是要从立法层面入手。

此外，以前**国内对网站数据造假的惩罚案例很少，惩罚并不算严格**，也是数据或者内容造假的重要原因之一。搜狐视频高级主管闵博认为，有关部门应加强对网络“爬虫”工具使用者的监管，对于恶意阻塞网络访问等滥用行为要严格查处，充分保障中小型互联网企业的网络空间安全使用权。

专家认为，未经对方允许从其他平台抓取数据并谋取商业利益的一般属于不正当竞争行为，由工商部门负责监管。但是因为这类行为通常比较隐蔽，工商部门一般需要有人举报并提供相应证据或线索后启动调查，因此需要借助技术手段更有效地进行打击。

诚信经营，遵法守法，互联网绝不是例外。正如北京知识产权法院在一起涉及爬虫技术的案件审判中所言，网络运营者应当遵循合法、正当、必要的原则，尽到网络运营者的管理义务。第三方应用开发者在收集、使用个人数据信息时，应当遵循诚实信用的原则及公认的商业道德。