你知道拖库，撞库和洗库么？

一看到这样的标题，我相信同学们是懵逼的。根据标题大概能猜出来是数据库相关的内容。但是心里还是有无数多个疑问！接下来我们从浅入深的说明这三种操作。

什么是拖库（因为谐音，也经常被称作“脱裤”）

拖库这个行为现在已经完全被黑产黑化了，说到拖库，马上就能联想到黑产。谁能想到，人家原来可是正经操作。

拖库本来是数据库领域的术语，指从数据库中导出数据。原本这么单纯的事，在不法黑客多次攻击事件之后，被用来指不法黑客入侵有价值的网络站点，把注册用户的资料数据库全部盗走的行为。

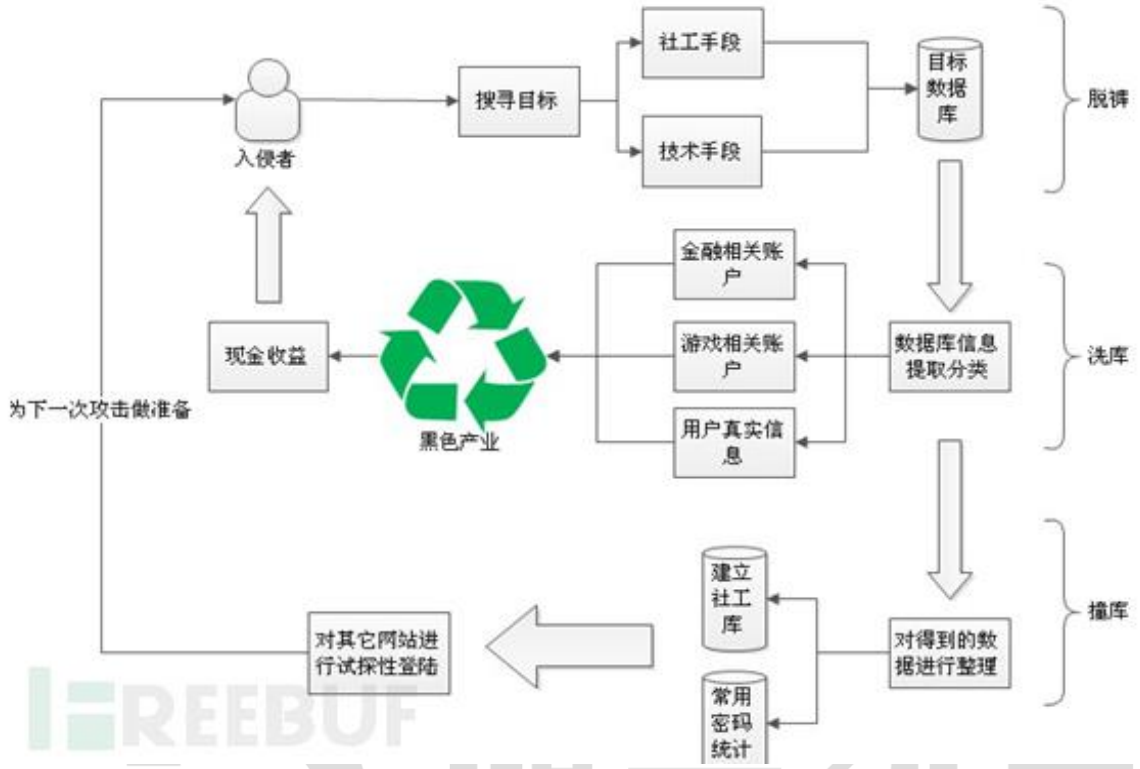
说到拖库，就不得不提他的手足兄弟 “撞库”，“洗库”

撞库与洗库

洗库：在取得大量的用户数据之后，黑客会通过一系列的技术手段和黑色产业链将有价值的用户数据变现，这通常也被称作“洗库”。

撞库：黑客将得到的数据在其它网站上进行尝试登陆，叫做”撞库“，因为很多用户喜欢使用统一的用户名密码，”撞库“也可以是黑客收获颇丰。

下图是黑客，在“脱裤”“洗库”“撞库”三个环节所进行的活动。



黑客怎样获取用户数据（拖库过程）

拖库的通常步骤为：

第一，不法黑客对目标网站进行扫描，查找其存在的漏洞，常见漏洞包括 SQL 注入、文件上传漏洞等；

第二，通过该漏洞在网站服务器上建立“后门(webshell)”，通过该后门获取服务器操作系统的权限；

第三，利用系统权限直接下载备份数据库，或查找数据库链接，将其导出到本地。

拖库的技术手段：

（1）远程下载数据库文件

这种拖库方式的利用主要是由于管理员缺乏安全意识，在做数据库备份或是为了方便数据转移，将数据库文件直接放到了 Web 目录下，而 web 目录是没有权限控制的，任何人都可以访问的；还有就是网站使用了一些开源程序，没有修改默认的数据库；其实黑客每天都会利用扫描工具对各大网站进行疯狂的扫描，如果你的备份的文件名落在黑客的字典里，就很容易被扫描到，从而被黑客下载到本地。

（2）利用 web 应用漏洞

随着开源项目的成熟发展，各种 web 开源应用，开源开发框架的出现，很多初创的公司为了减少开发成本，都会直接引入了那些开源的应用，但却并不会关心其后续的安全性，而黑客们在知道目标代码后，却会对其进行深入的分析和研究，当高危的零日漏洞发现时，这些网站就会遭到拖库的危险。

（3）利用 web 服务器漏洞

Web 安全实际上是 Web 应用和 Web 服务器安全的结合体；而 Web 服务器的安全则是由 Web 容器和系统安全两部分组成，系统安全通常会通过外加防火墙和屏蔽对外服务端口进行处理，但 Web 容器却是必须对外开放，因此如果 Web 容器爆出漏洞的时候，网站也会遭到拖库的危险。

社会工程学方面大概有如下几种：

（1）水坑攻击

黑客会利用软件或系统漏洞，在特定的网站上进行挂马，如果网站管理员在维护系统的时候不小心访问到这些网站，在没有打补丁的前提下，就会被植入木马，也会引发后续的拖库风险。

（2）邮件钓鱼

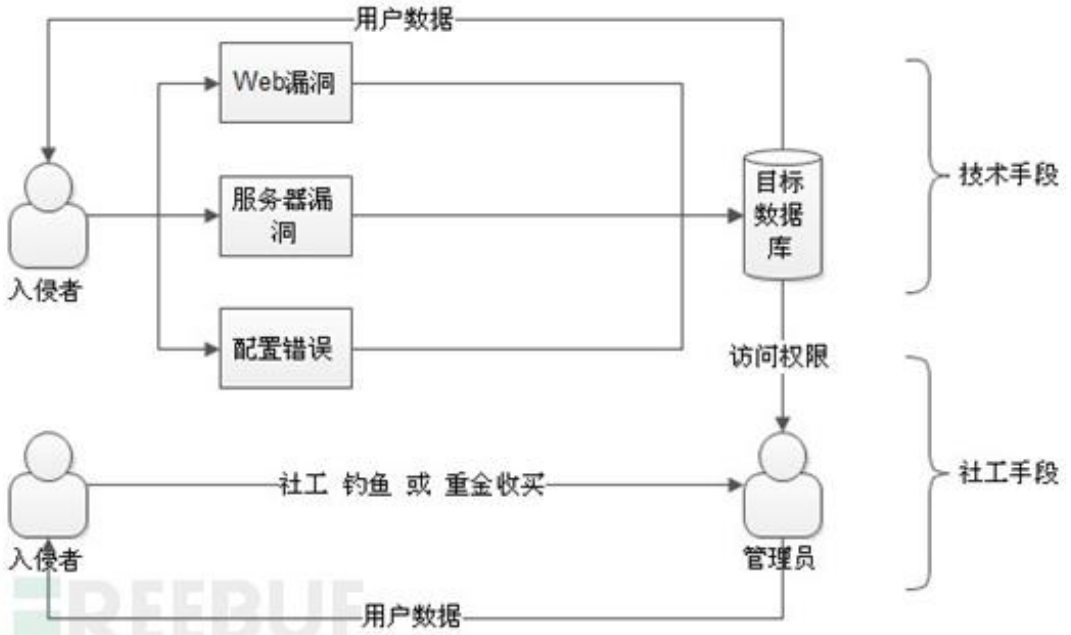
黑客会利用一些免杀的木马，并将其和一些管理员感兴趣的信息绑定，然后通过邮件发送给管理员，而当网站管理员下载运行后，也会导致服务器植入木马，引发后续的拖库风险。

（3）社工管理员

对目标网站的管理员进行社会工程学手段，获取到一些敏感后台的用户名和密码。从而引发的后续拖库。

（4）XSS 劫持

有时黑客也会为了获取某一些网站的帐号信息，他们会利用网站钓鱼的手段去欺骗用户主动输入，但这种方式只能获取部分帐号的真实信息，并没有入侵服务器。



如何避免“脱裤”？

1.从数据库角度防范拖库

数据库密码存储的防拖库设计

目前，大多数网站对数据库中的密码信息是加密存储的。常见的有MD5 加密，理论上说这种方式是不可逆的，但是这种方式仍然是不安全的，只要枚举出所有的常用密码，做成一个索引表，就可以推出来原始密码，这张索引表也被叫做“彩虹表”。

面对以上风险，主流网站后端数据库的密码存储都在 MD5 的基础上进行加盐；所以不再只保存加密过的口令，而是先将口令和随机数连接起来然后一同加密，加密后的结果放在口令文件中。

随着技术的发展，出现了强认证技术，即二次认证密码：除了对固有密码进行加密存储外，系统还自身生成一个加密密钥。已广泛应用在金融行业及支付行业，我们使用比较多的有手机短信验证码、动态令牌、Ukey、密钥文件等形式。

数据库数据交互防拖库设计

根据自身业务特征、自身 IT 技术实力，在前端应用与后端数据库的交互过程中，建议设计统一查询接口，便于管理和监控，设置黑白名单策略。

在对数据库管理与维护的过程中，可以使用专业的运维防护工具，如数据库安全网关、堡垒机等。

对数据库进行全方位的监测，包括访问情况、数据交互情况、风险操作情况、网络流量等信息，一旦发现危险操作可及时处置。

2.用户角度自身防范

1、分级管理密码，重要帐号（如常用邮箱、网上支付、聊天帐号等）单独设置密码；

2、定期修改密码，可有效避免网站数据库泄露影响到自身帐号；

3、工作邮箱不用于注册网络帐号，以免密码泄露后危及企业信息安全；

4、不让电脑“自动保存密码”，不随意在第三方网站输入帐号和密码。即便是个人电脑，也要定期在所有已登录站点手动强制注销进行安全退出；

用户数据与黑色产业（洗库操作）：

随着地下产业链日渐成熟，用户数据可以被迅速地转变成现金。

（1）用户账号中的虚拟货币，游戏账号，装备，都可以通过交易的方式变现，也就是俗称的“盗号”。

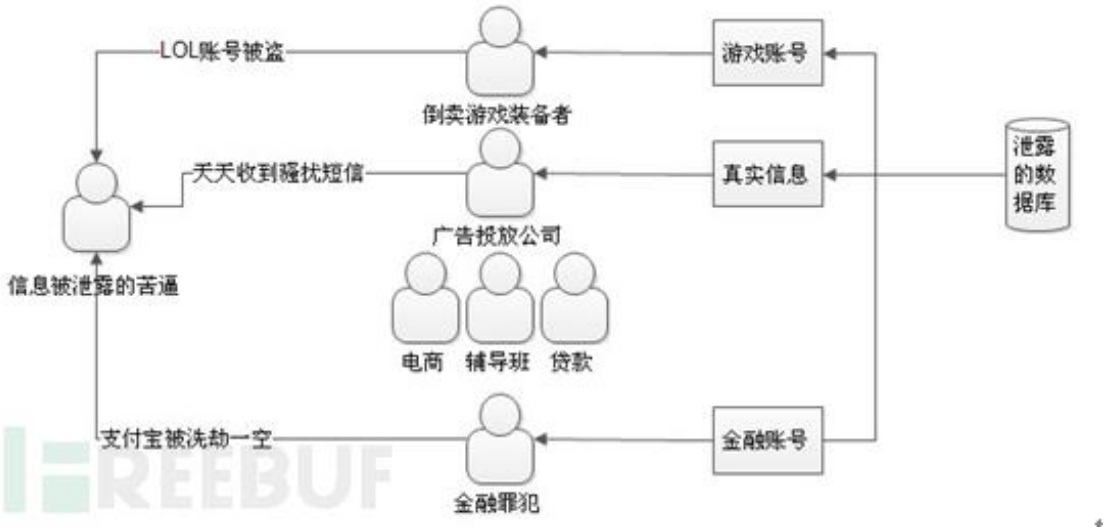
（2）金融类账号比如，支付宝，网银，信用卡，股票的账号和密码，则可以用来进行金融犯罪和诈骗。

（3）最后一些可归类的用户信息，如学生，打工者，老板等，多用于发送广告，垃圾短信，电商营销。也有专门的广告投放公司，花钱购买这些分门别类的信息

快速收益和高回报也让越来越多的黑客铤而走险。（刑法里非法入侵计算机系统罪会被判处三年到七年有期徒刑）

而对于，信息被泄露的受害者，根据泄露信息的种类不同，生活也会受到不同程度的影响。

如下图：　如果你的多种网站和服务的用户名密码相同，那可能会蒙受更大的损失。对于洗库操作没有好的避免方式，应该从防止拖库开始，进行避免。



常见的撞库操作

撞库是黑客通过收集互联网已泄露的用户和密码信息，生成对于的字典表，尝试批量登录其他网站后，得到一系列可以登录的用户。因为很多用户在不同网站使用的账号密码大多是相同的，因此黑客可以通过获取用户在 A 网站的账户从而尝试登录 B 网站。

最常见的三种撞库方法：

第一种：用 n 个密码字典撞 m 个账号，这个的表象是，一个账号在某个较短的时间内，可能会有多次密码尝试。所以，可以在账号层加限制措施，比如：**一天内，一个账号，密码错误次数超过 5 次时，1 天之内禁止登陆**（或者校验手机短信/密保问题之后才能登陆）。

第二种：用几个密码撞 n 个账号，这个的表象是，密码出现的频率会非常高，所以，可以**统计一段时间内每个密码的错误次数，超过一定阈值时，这个密码在一段时间内禁止登录**（或者校验手机短信/密保问题之后才能登陆）。

第三种：用 n 组一一对应的账号密码来再撞库，这种情况的撞库单纯从账号、密码的维度来看，不会有明显的异常。

如何防范撞库操作

1、**IP 封禁**，如果一段时间内，单个 IP 地址，密码错误次数超过阈值，则禁止这个 IP 一段时间再登录（或者校验手机短信/密保问题之后才能登陆）。不过，如大家所说，现在代理 IP 相当廉价，从 IP 层面来封禁基本上没啥作用。

2、**建立 IP 画像库**，对代理 IP、IDC IP 等高危的 IP 直接禁止登陆（或者校验手机短信/密保问题之后才能登陆）。自己建立 IP 画像库成本可能会有点高，可以考虑采购安全厂商的类似服务。

3、**现在比较火的行为验证码**，比如：拖条、点选、拼图等各种花样的验证码。只是说，如果之前登录不需要验证码，现在要加上一个验证码，估计要和产品撕逼。一般来说最后为了后期的运营，产品也会同意加上验证码。

4、**从设备层面来识别和封禁**，通过在客户端植入 sdk，收集用户端的设备信息，从设备层面来做高频策略，或者，直接识别出非正常的设备，然后对设备进行封杀。

5、**从行为层面来识别和封禁**，和上面一条一样，通过客户端植入 sdk，收集用户在登录页面的交互行为，通过机器学习、大数据建模，训练出正常用户、异常用户的行为模型，在交互行为层面，将撞库的行为识别出来。这个需要有预先训练好的行为模型，现在机器学习那么好，不说大家也都知道，自己训练一个模型肯定需要很多标注数据，这也就意味着成本。所以，还是建议寻找安全厂商还做，毕竟专业的人做专业的事，靠谱！