数据安全发展态势及相关技术

文 / 戴林

2017 年 6 月 1 日正式实施的《网络安全法》第十条要求“建设、运营网络或者通过网络提供服务，应当采取技术措施和其他必要措施，维护网络数据的完整性、保密性和可用性。”我们可以把这个要求看成是当前数据安全的正式定义。

广义的数据安全技术是指一切能够直接、间接地保障数据的完整性、保密性、可用性的技术。这包含的范围非常广，比如传统的防火墙、入侵检测、病毒查杀、数据加密等，都可以纳入这个范畴。正因为如此，很多传统的安全厂家都给自己贴上“数据安全厂家”的标签。

而狭义的数据安全技术是指直接围绕数据的安全防护技术，主要指数据的访问审计、访问控制、加密、脱敏等方面。而这里的数据，则可以粗略地分为两类：一类是非结构化的数据，例如图片、文件、图纸等；另一类是结构化的数据，主要存储于数据库中。当然非结构化的数据很大一部分也存储于数据库中，尤其是现在的各种 NoSQL 数据库，就是专门针对非结构化数据设计的。而相应的数据安全技术，也可以粗略地划分为针对非结构化数据的安全技术和针对结构化数据的安全技术。

本文聚焦于狭义的数据安全。

对于非结构化的数据安全，主要采用数据泄露防护（Data leakage prevention, DLP）技术。DLP 技术发展相对成熟，国外比较具有代表性的有 Symantec 的 DLP 产品，国内也有不少类似产品。而对于结构化的数据安全技术，国外发展比国内早 5~10 年。个别产品，如数据库审计、数据库防火墙，以及数据库脱敏，现在国外进入产品和市场的成熟期，代表性厂家有 Imperva、IBM Guardium、Infomatica 等，而国内企业目前勉强有产品能够进行替代，实际差距还比较大。在针对云环境和大数据环境的安全方面，国内刚刚起步。以下逐个盘点。

数据泄露防护 DLP

调查显示，在文档类泄密事件中，97% 都是因内部员工有意或无意泄露而造成。其主要原因为核心数据大多以文件为载体，零散分布在员工电脑及移动介质中，且以明文存储，不受管控。文档的使用者可任意编辑、拷贝、转发、打印等，文档处在“裸奔”状态，存在巨大的安全隐患。数据泄露防护（DLP）基于文档加密，进而控制其解密权限，从根源上防止数据外泄。

DLP 的核心技术在于如下几点：

1. 动态透明加解密，用户无感知，在不影响用户办公习惯的前提下，有效控制使用者对文档的读取、存储、复制、输出的权限，防止数据泄密。

2. 文档分级管控，根据文档流转范围，对文档进行多级别分级管理，确保文档接收者只能做权限范围内的操作。

3. 文档外发管理，对受控的外发文件进行加密和权限设定，防止第三方泄密。

4. 系统集成拓展，可与文档安全网关、邮件安全网关等系统整合，实现对单位应用系统的安全集成及对核心数据载入载出的安全保护。

由于国外产品 Symantec 曾经被审查出后门程序，以及国内密码管理政策原因，目前国内的 DLP 产品主要采用国内自主研发为主，并且由国内公司研发生产，目前通过应用层及驱动层加密相互配合，可实现任意文档类型的加密处理，并且没有文档大小及类型的限制，文档处理效率几乎不受影响，技术已基本成熟。

数据备份与容灾

需要确保数据备份和容灾系统通过建立数据的备份以及远程的容灾备份来确保在发生灾难性事件时，数据能够被正常地恢复，从而提升数据的可用性。

数据容灾备份的关键技术在于：

1. 数据变化的捕捉，将差异变化以最小的代价传送到备份端。

2. 恢复技术，需要在最短的时间甚至是零时间内将备份数据恢复到生产系统中。

目前数据与容灾的市场和技术都相对成熟，国内厂家较多，产品的可选择余地较大，基本可以完全替代国外产品。

数据库审计 / 防火墙

数据库审计是最基础的数据库安全手段。由于数据库是个“黑盒子”，对来自内网、外网的用户和系统对核心数据的访问情况，尤其是违规访问情况缺乏可视化。数据库审计通过分析访问数据库的网络流量，对数据的访问情况进行展示，并进一步地识别敏感数据的窃取和破坏行为，比如对 SQL 注入攻击、后门程序等进行识别。而数据库防火墙则更进一步的，设置对核心数据的访问规则，阻止来自内网用户的越权访问和误操作。并且这种访问规则是独立于数据库系统自身的访问控制。

数据库审计 / 防火墙的核心技术在于如下几点：1. 高效的数据包获取、分析和转发技术；2. 完整、准确的数据库协议解析和 SQL 协议解析；3. 灵活有效的访问控制规则系统；4. 自动学习能力，能够自主地学习用户对数据的访问模型，并基于该模型进行访问控制；5. 高效的日志存储和查询性能；6. 灵活的部署方式，能够部署于多种应用环境。

国外知名的数据库审计 / 防火墙厂家有 Imperva 和 IBM Guardium，它们的产品前几年在国内大型 IT 系统中部署较多。现在国内一些公司的产品在界面、功能、性能等方面逐步接近国外产品，基本能够替代。

数据库加密

敏感数据在数据库中明文存储，会使得存储文件、磁盘或者备份文件等被非法复制时导致数据泄露，而且商用数据库还面临着管理员权限过大，导致权利和责任的不统一。也就是说数据管理员（DBA）不应该有查看或者删除所有敏感数据的权限，但是他实际上却拥有这种权限。这也将导致数据的泄漏，尤其是在 DBA 权限被泄露的情况下。数据库加密就是对敏感数据字段进行选择性的加密，并建立独立于数据库的访问控制规则，从而弥补上述风险。

数据库加密的核心技术在于如下几点：1. 对应用透明，包括增删改查四种操作，以及主键、外键、约束等特性，修改表定义等操作；2. 密文索引，确保加密后数据的查询性能不受实质影响，也就是返回首条记录的时间没有本质的延长；3. 与国密算法的集成。

受政策、价格等原因的影响，数据库加密产品主要是国内创业公司生产的产品。由于密文索引的实现依赖于数据库开放的自定义索引接口，目前主流数据库中，仅有 Oracle 数据库提供这种接口，所以到现在为止，市场上真正成熟的数据库加密，只有针对 Oracle 的产品。目前国内数据安全创业公司在开发针对 MySQL 的数据库加密产品，主要目标是各个共有云平台上大量的 MySQL 用户。技术路线有修改存储引擎和加密网关两种。修改存储引擎方式比较简单，对 SQL 的通用性好，但是只适合开源数据库产品；而网关型加密产品对数据库的通用性较好，但是对数据库的某些特性支持起来比较困难。

数据库脱敏

数据脱敏技术分为动态脱敏和静态脱敏。静态脱敏针对的是在开发、测试过程中使用真实敏感数据可能会导致的数据泄密风险。静态脱敏类似于 ETL，对真实数据进行定时、批量的抽取以及脱敏转换，从而提供准确真实的数据。而动态脱敏针对内部运维人员、外包人员在系统运维过程中，接触真实敏感数据，容易导致泄密的风险，以及应用系统直接访问敏感数据，获取真实数据内容，容易导致泄密的风险。动态脱敏系统部署于数据之前，通过改写访问数据库的语句，从源头上选择性地对敏感数据进行脱敏，并可以控制对敏感数据的访问总量。

数据库静态脱敏的核心技术在于如下几点：1. 高速的数据抽取和装载技术；2. 灵活的脱敏规则；3. 敏感数据发现和随机数据生成能力；4. 脱敏后关联关系的保持。

数据库动态脱敏的核心技术在于如下几点：1. 高效的数据包获取、分析和转发技术；2. 完整准确的数据库协议解析和 SQL 协议解析；3. 灵活有效的脱敏和访问控制规则系统；4. 三层脱敏和访问控制。

国外数据库脱敏的代表厂商有 Informatica 和 IBM OPTIM。而国内的脱敏市场和产品都是近两年才发展起来的，整体落后。但是国内的数据库脱敏技术发展很快，目前静态脱敏产品有一定选择余地，基本可以替代国外产品。但是动态脱敏由于技术难度更大，产品仍然很少，选择余地不大，但是仍然基本可以替代国外产品。

云环境数据安全

在云端，数据所面临的威胁被进一步放大。除了遭受与传统环境相同的安全威胁以外，由于云运营商的存在，数据还遭受“上帝之手”的威胁。以数据库为例来说，在云端，数据库的租户对数据库的可控性是很低的，甚至不能登录到数据库所在的 OS 进行管理。而云运营商却拥有对数据库以及其服务器的所有权限。云运营商完全可以在租户毫无察觉的情况下进入数据库系统，或者进入数据库服务器所在的虚拟机。也就是说，云数据库租户在数据库中的数据，对云运营商来说，几乎是完全开放的。这极大地增加了存储在云端数据库中具有商业价值的数据被泄露的风险。

云端数据对安全的技术需求，与线下是一致的，也需要借助于审计、访问控制、加密、脱敏等技术的保护。但是由于云管理员的存在，云端数据对加密的要求是第一位的，也就是说，首先要确保数据的保密性，这是区别于线下数据安全的最显著的特征。

另外，技术之外的一个要求就是客观中立性。对于云端的数据安全防护，最好应该是来自第三方。所谓“第三方”，就是指这种安全措施，不是由云运营商提供的，而是由其他独立厂商提供的，以避免管理上和技术上的后门。

国外云端数据的安全厂家比较有代表性的如 CiperCloud 和 Skyhigh Networks，国内云端数据安全方案刚刚起步，有一些审计类的产品开始部署，但是加密类的产品，还在研发阶段。CEN

（作者单位为北京理工大学）

大数据平台的数据安全

在流行的大数据平台 Hadoop、Cloudera 和 Splunk 中，数据存储和处理的方式发生了很大的变化。既可以采用传统关系型数据库系统，又可以采用新型的 NoSQL 数据库，如 HBASE，Mongodb，Cassandra，Hive 等。当采用新型 NoSQL 数据库时，数据安全面临新的问题。目前国外针对 Hadoop 和 Cloudera 环境有数据库审计、数据库防火墙等产品。但是国内在此领域只有极少数公司在进行试探性的研发，目前尚未有相对成熟的产品上市。

《网络安全法》对数据安全的相关要求

《网络安全法》是我国关于网络空间安全的首部法律，是一部基本法。在《网络安全法》中，有大量篇幅的内容是与数据安全相关的，涉及到技术和管理两个方面的内容。概括起来，《网络安全法》中有关网络数据安全的技术性要求有如下几点：

1. 对数据访问日志进行审计，且日志留存时间不低于 6 个月（第 21 条）；
2. 对数据进行分类，将敏感数据与普通数据区别化处理（第 21 条）；
3. 对重要数据进行备份，容灾（第 21、34 条）；
4. 对重要数据进行加密（第 31 条）；
5. 对个人信息进行脱敏（第 42 条）。