绿盟数据安全解决方案
白皮书

绿盟科技

2019 年 2 月

目　　录

序：

人类经历了三次数据量的跃升，Web 1.0 时代以门户网站为主，Web 2.0 时代用户原创内容带来“数据爆炸”，物联网时代数据上 TB、PB 级，进入“大数据时代”，IDC 估测，数据以每年 50%的速度增长。

这是一个互联，且不断保持“在线”的社会，数据只有被共享才更有价值。

当今数据的来源复杂而多样，云计算、大数据、物联网、移动互联网、车联网、手机、平板电脑、个人电脑（PC）以及遍布地球各个角落的各种各样的传感器，无一不是数据来源或者承载的方式。

数据是未来最大的资产，用好数据不仅可以提高企业自己的产品和服务，也可以攫取大量利润。一旦没有守好数据，那么很有可能成为下一个负面信息的主角。

数据安全已经成为全世界瞩目的焦点问题，连续三年蝉联 RSA 热词榜冠军。

1 数据安全风险分析

移动互联网的普及使得人们越来越多的在网络上留下信息，这些信息如果被分析和利用，将对个人隐私和安全形成极大的威胁，同时海量数据也增加了信息保护的难度。

历年来，数据泄露事件愈演愈烈，透过事件我们看到了问题的本质：

- **对数据的违规使用**

  非法收集：

  外部：漏洞攻击、木马注入、弱配置、APT

  内部：越权盗窃、离职

  数据滥用：

诱导、贩卖、敲诈

- **数据泄露带来的风险：**

  访问：认证、权限

  共享：业务（门户、调用测试）、人员交互

  外发：跨区、第三方（网络、邮件）

  外带：出差、回家

2 政策上的措施

针对不断涌现的数据泄露问题，数据和隐私保护政策陆续出台：

- 我国于 2017 年 6 月 1 日正式施行《中华人民共和国网络安全法》，规定了公民使用网络服务需要实名认证，任何网络侵入、干扰和窃取网络数据都是违法的，个人信息安全得到真正的法律保护，从此确立了公民个人信息保护的基本法律制度，促进经济社会信息化健康发展。
- 我国《网络安全等级保护条例》提出对信息进行收集、存储、传输、交换、处理的系统进行不同等级的保护要求。对定级不准确不合理的网络运营者，应准确履行自己的网络安全义务，工作不到位的网络运营者主要负责人以及网络安全相关负责人将受到响应的处罚。
- 我国《关键信息基础设施安全保护条例》 提出对保护范围内的单位运行、管理的网络设施和信息系统，一旦遭到破坏、丧失功能或者数据泄露，可能严重危害国家安全、国计民生、公共利益的，都应受到网络安全法的处罚，为企业带来了合规挑战。
- 我国于 2018 年 5 月 1 日正式实施《个人信息安全规范》，规范个人信息控制者在收集、保存、使用、共享、转让、公开披露等信息处理环节中的相关行为，进一步强调了个人敏感信息被泄露、非法提供或滥用可能危害人身、财产安全，致使个人名誉、身心健康受到损害或歧视性待遇等严重后果，遏制个人信息非法收集、滥用、泄漏等乱象，最大程度地保障个人的合法权益和社会公共利益。
- 欧盟于 2018 年 5 月 25 日正式施行《通用数据保护条例》，简称 GDPR，被称为史上最严数据保护法，其最高惩罚代价为暂停使用个人数据。

3 数据安全防护思路

3.1数据安全防护目标

绿盟科技为数据安全设计了全面可信的防御体系，有效保护数据在全生命周期过程中的安全，达到合法采集、合理利用、静态可知、动态可控的防护目标。

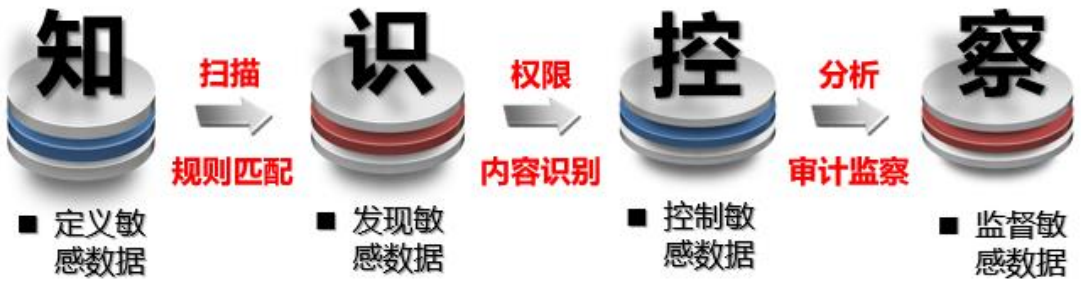
- **合法采集：**利用大数据分拣技术，使企业在法律约束范围内合法采集敏感数据；

- **合理利用：**通过建立数据模型，以及对数据的敏感级别进行划分，设立不同的访问层级，在数据被开发利用前做好防护措施，杜绝非法滥用；
- **静态可知：**对存储中的静态数据进行扫描发现，并展示数据的分布；
- **动态可控：**对流动的数据进行监控，防止数据在交互、共享中有意无意的泄露。

3.2 数据安全治理设计思路

数据安全就是对数据的安全治理，是从政策到数据的全生命周期的监察与保护。

绿盟科技结合客户的需求，以及对实际环境的调研了解，总结出了一套完整又科学的数据安全治理方法，及“知”、“识”、“控”、“察”。



- **知：**分析政策法规、梳理业务及人员对数据的使用规范，定义敏感数据；
- **识：**根据定义好的敏感数据，利用工具对全网进行敏感数据扫描发现，对发现的数据进行数据定位、数据分类、数据分级。
- **控：**根据敏感数据的级别，设定数据在全生命周期中的可用范围，利用规范和工具对数据进行细粒度的权限管控。
- **察：**对数据进行监督监察，保障数据在可控范围内正常使用的同时，也对非法的数据行为进行了记录，为事后取证留下了清晰准确的日志信息。

将数据安全治理方法“知”、“识”、“控”、“察”应用于实际项目中，利用咨询服务发现数据风险，通过产品落地实现对数据的可视化监控、风险点排除，及时预警、及时阻止对数据的非法使用行为，最后对数据进行持续运营服务，让数据始终处于被监控的安全状态，当有新的业务上线时，可根据此数据治理方法快速的实现新数据的安全监控。

4 数据安全解决方案

绿盟科技针对数据安全提出了完整的解决方案，包括数据梳理、运维数据监管、业务数据监管、办公数据监管，以及数据的可视化，全面对数据在各种场景中的全生命周期安全进行了阐述。

4.1 数据梳理与风险评估

- **数据系统础调研：**对关键数据及用户敏感信息的数据系统进行初步调研；
- **关键数据定义及分级：**对各系统中关键数据及用户敏感信息等数据进行定义，并对关键数据、用户敏感信息按重要性进行分类、分级；
- **数据生命周期梳理：**从数据的采集、传输、存储、使用、共享、销毁等环节入手，覆盖人员、设备、系统三要素，梳理各系统在各数据生命周期环节中相关网络关键数据详情；
- **数据安全合规性检测：**从数据安全制度管理、数据安全运营管理、数据安全生命周期管理等纬度梳理数据合规性安全现状；
- **业务安全合规性检测：**对网络数据与用户数据集中的业务应用进行金库模式安全性、批量数据违规获取、身份认证安全性、敏感信息模糊化、系统授权管理、防撞库攻击安全性等多方面业务安全合规性检测；

- **数据安全管理制度建设：**在安全管理制度的制定、落实和审查的三个环节上形成可持续完善的机制；
- **网络数据泄露防护：**通过采用 DLP 等网络数据泄漏技术监测手段，及人工主动干预方式，对疑似泄漏渠道进行分析验证，对风险点进行识别、发现和处理；
- **数据保护审计策略制定：**对系统制定相关数据保护审计制度及详细审计策略，包括网络数据安全保护审计、业务符合性审计等；
- **数据保护安全审计服务：**从数据的采集、传输、存储、共享、销毁等环节入手，建立、健全相关审计机制，并定期开展专项审计工作；
- **安全检查支撑服务：**针对上级单位安全检查、重大活动期间专项保障等期间提供支撑服务。

数据梳理与风险评估服务企业带来的价值主要体现在：

- 形成全局数据分布情况，奠定分级分类管理基础
- 全方位了解数据安全现状与所面临的安全威胁
- 提供专业整改建议，促进系统整改
- 建立标准化、规范化、专业化数据安全管理体系
- 提升数据安全技术防护水平
- 针对性地、有序地进行各项日常数据安全工作和开展数据安全建设

4.2 运维数据安全防护

随着信息化的发展，企事业单位 IT 系统不断发展，网络规模迅速扩大、设备数量激增，建设重点逐步从网络平台建设，转向以深化应用、提升效益为特征的运行维护阶段，IT 系统运维与安全管理正逐渐走向融合。信息系统的安全运行直接关系企业效益，构建一个强健的 IT 运维安全管理体系对企业信息化的发展至关重要，对运维的安全性也提出了更高要求。

日常工作监管

通过对运维数据安全防护可以对运维人员和合作伙伴的日常操作情况做防护和记录，方便监管。

法律法规遵循

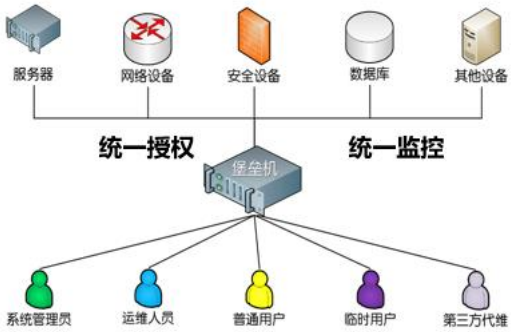
许多企业和单位需要满足国家或者行业监管部门的法律法规要求(如等级保护、企业内控制度等)。

事后追溯

通过事后审计信息可以对发生的安全事件进行追溯，保留一份不能修改不可删除的“证据”。

4.2.1运维统一监管

堡垒机做为专业的运维监管系统，提供了先进的运维安全管控与审计能力，是运维数据防护的第一道防线，其目标是帮助企业转变传统 IT 安全运维被动响应的模式，建立面向用户的集中、主动的运维安全管控模式，降低人为安全风险，对运维数据的访问行为实现全面的审计，满足合规要求，保障企业效益。



堡垒机通过逻辑上将人与目标设备分离，建立“人->主账号（堡垒机用户账号）->授权->从账号（目标设备账号）->目标设备”的管理模式；在此模式下，通过基于唯一身份标识的集中账号与访问控制策略 ，与各服务器、网络设备、安全设备、数据库服务器等无缝连接，实现集中精细化运维操作管控与审计。

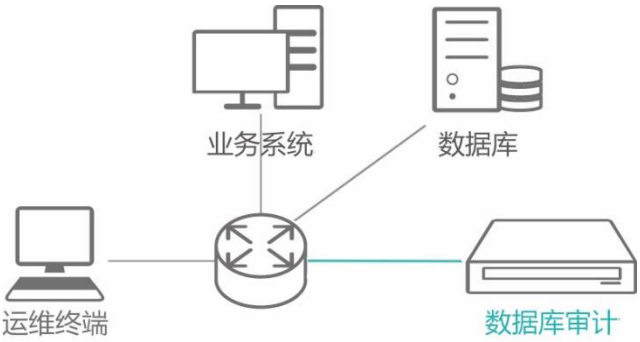
堡垒机为企业带来的价值主要体现在：

- 可帮助企业建立面向用户的集中、有序、主动的运维安全管控平台；
- 通过基于唯一身份标识的集中账号与访问控制策略，与各服务器、网络设备等无缝连接，实现集中精细化运维操作管控与审计；
- 对运维数据实时监控，日志+录屏双重审计，保障运维数据行为安全；
- 降低人为安全风险，避免安全损失，满足合规要求，保障企业效益。

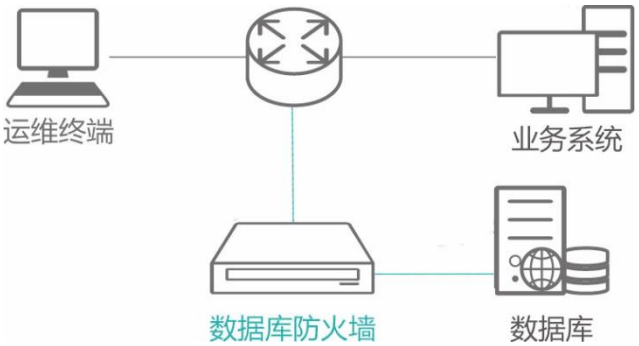
4.2.2数据库审计与防护

近年来，随着数据篡改、泄密等事件的频繁爆发，行业用户已经普遍意识到数据库访问疏于监管所带来的巨大危害，数据库审计与防护产品受到空前关注。其中就包括数据库审计和数据库防火墙。

数据库审计是通过对数据库网络流量的采集，利用数据库协议解析与还原技术，实现对数据库所有访问行为的监控和审计、对其中的危险操作进行多种方式的告警、对数据库访问行为进行多维度的统计并进行图形化展现。

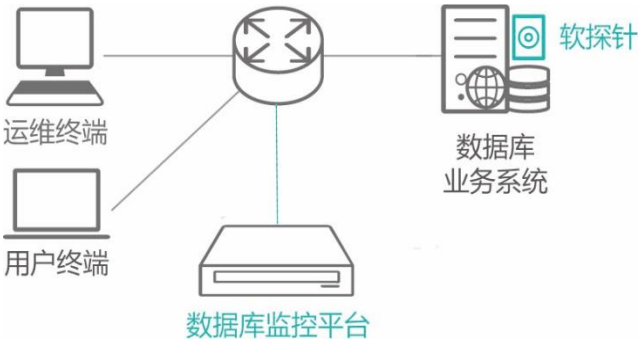


数据库防火墙是利用数据库协议分析与控制技术，基于主动防御机制，实现数据库的访问行为控制、危险操作阻断、可疑行为审计。



将数据库审计与数据库防火墙结合使用，就实现了具有集数据库 IPS、IDS 和审计功能为一体的综合安全防护能力。

现如今正处于一种多环境数据共享的时代，如传统环境、云环境、大数据环境等，因此对数据库的审计与防护产品就需要做适配变化，传统环境依然采用传统硬件部署，云环境应满足虚拟化要求实现软件化部署或者探针模式部署，大数据环境一般都应满足分布式部署的要求，最终可以通过集中管控形成统一监管的模式。



数据库审计与防护为客户价值带来的价值包括：

- 防止外部黑客攻击；
- 防止内部高危操作；
- 防止敏感数据泄漏；
- 审计追踪非法行为。

4.2.3大数据安全运维

大数据的思想及其初步应用已经惠及人们的日常生活，与大数据相互依存的云计算技术、物联网、智慧城市等新的应用模式同时印证了其在信息化时代的重要地位。随着数据的价值越来越重要，大数据的安全稳定也逐渐被重视，在大数据时代，无论对于数据本身的保护，还是对与由数据而演变的一些信息的安全，都对大数据环境提出了更高的要求。虽然大数据安全与大数据业务是相对应的，但对于业务和环境的维护将更为重要，大数据安全运维将包含风险检查、控制防护、审计监控三个方面。

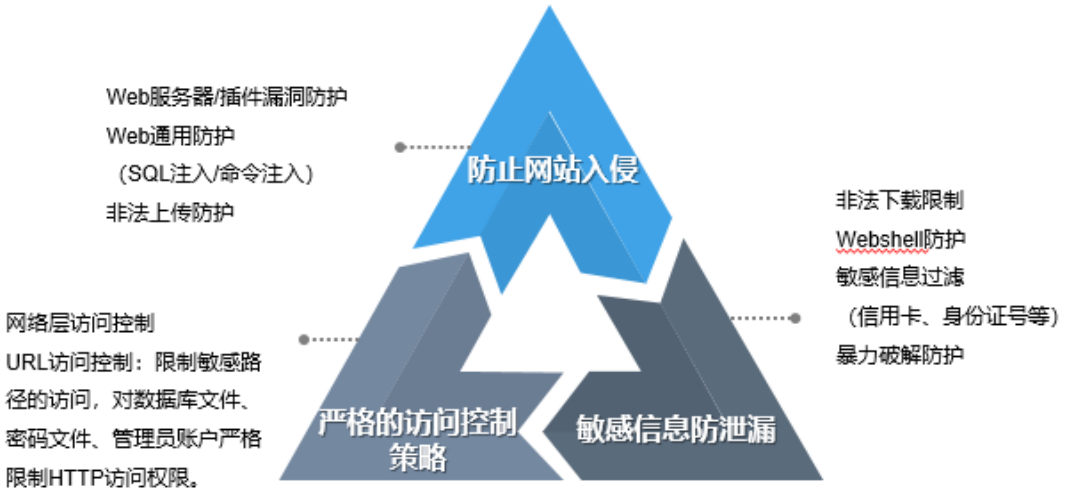
- 风险检查：可以发现大数据组件漏洞与配置威胁、定位大数据中的敏感数据，让运维人员对大数据环境中的数据风险了如指掌。
- 控制防护：通过实时获取请求者、环境和被访问数据三要素的属性信息，触发相应的访问控制策略，从而实现访问控制的动态管理，降低企业的管理和运营成本。
- 审计监控：针对大数据环境中的各个组件的数据调用与交互，准确的对数据进行审计、回溯、风险控制、职权分离、操作过程回话等功能，同时生成访问日志，为运维人员提供可分析的有效数据。

大数据安全运维为客户价值带来的价值包括：

- 大数据组件风险全面监控；
- 准确定位敏感数据的安全风险，快速预警；
- 敏感数据访问动态管理，权限明确，风险可控；
- 日志全面完整，为事后回溯提供有利数据支撑。

4.3 业务数据安全防护

随着信息技术日新月异的发展，近些年来，企业利用计算机网络技术与各重要业务系统相结合，可以实现无纸办公。有效地提高了工作效率，如外部门户网站系统、内部网站系统、办公自动化系统等。然而信息化技术给我们带来便利的同时，各种网络与信息系统安全问题也逐渐暴露出来，业务数据泄露事件频发。因此要对业务系统进行安全防护，保障业务系统中的静态数据、动态数据的安全。



4.3.1WEB 应用数据防护

Web 应用防火墙（简称 WAF）将客户资产作为组织 Web 安全解决方案的依据，用黑、白名单机制相结合的完整防护体系，通过精细的配置将多种 Web 安全检测方法连结，并整合成熟的 DDoS 攻击抵御机制，能够抵御各类 Web 安全威胁和拒绝服务攻击，并以较低的运营成本为各种机构提供透明在线部署、路由旁路部署、镜像部署和云部署，能方便快捷的部署上线，保卫您的 Web 应用数据免遭当前和未来的安全威胁。

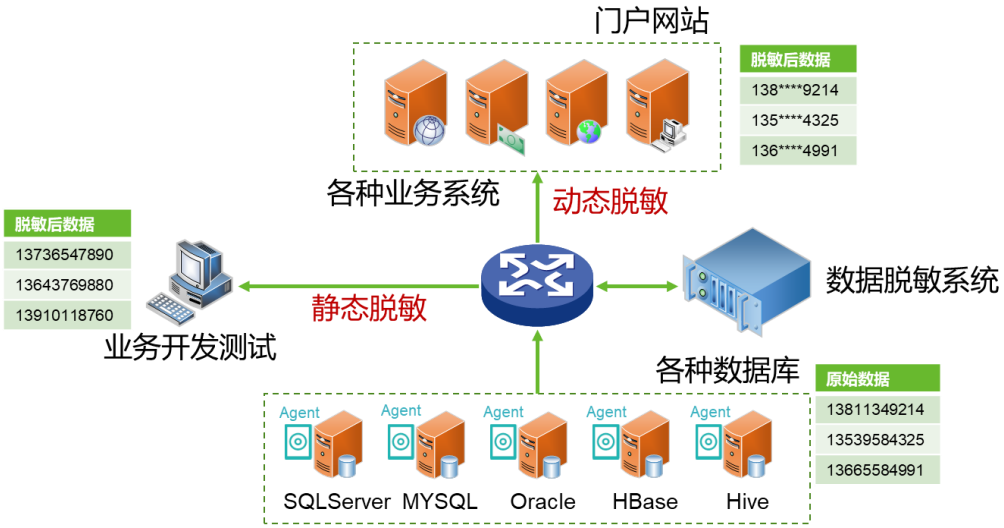
根据典型的业务数据泄露事件可分为五种防护场景：

- **网站访问控制：**不仅可以达到权限控制的效果，还可以做到误报纠正；
- **网页篡改在线防护：**提供事中防护以及事后补偿的在线防护；
- **敏感信息泄漏防护：**可以识别并防护敏感数据泄漏，满足合规与审计要求；
- **虚拟站点防护：**能对一个 IP 对应多个不同域名的虚拟站点场景进行防护。

4.3.2业务数据脱敏

数据脱敏技术可实现自动化发现源数据中的敏感数据，并对敏感数据按需进行漂白、变形、遮盖等处理，避免敏感信息泄露。同时又能保证脱敏后的输出数据能够保持数据的一致性和业务的关联性。

数据脱敏技术可以满足为开发环境、测试环境、培训环境等提供脱敏后的生产数据，也可于为数据交易、数据交换、数据分析等第三方数据应用场景提供适用的敏感信息泄露防护作用。



以防止数据泄漏为核心点，最大限度的保证脱敏后数据的特征、结构、逻辑及各类数据间的关联性、连续性、业务性。通过对客户业务数据的深度了解，保证整个业务系统正常运行。

4.4 办公数据安全防护

当前无纸化办公已经普遍应用，办公环境中每位员工都拥有独立的办公终端，企业的研发源代码，财务、政券信息，运营资料，人资等敏感数据以不同的文件形式存在于每个人的终端电脑及存储服务器中，员工通过办公终端接入网络，连接到各种业务系统、各类服务器进行数据的使用与交换，很难对敏感数据进行准确的定位和分析发现。

数据在办公环境中的类型多种多样，大体可分为结构化数据、半结构化数据与非结构化数据，还可按形态分为存储中的静态数据和使用中的动态数据。

对数据的操作包括：创建、复制、编辑、剪切、截屏、保存、另存为、删除等。

数据传输途径包括：USB、打印、刻录、共享、网络上传、邮件、论坛、微博、聊天工具、网盘，还有其他外设。

当前无论是终端，网络还是管理方面，都对敏感数据的泄漏存在有很大的风险。因此，需要采取相应的措施来消除这些威胁，降低整体安全风险，确保内部办公环境下的数据安全，具体防护方案可以归纳为以下几个方面：

4.4.1数据外发途径防护

基于敏感信息识别技术的防护对办公终端的数据泄露途径进行全面的监控。

可管理通过终端泄漏敏感数据的多个途径，如打印，U 盘拷贝，硬盘对拷，蓝牙/红外发送文件，光盘刻录，文件共享，IM 发送聊天内容/传送附件，如 QQ,RTX 等，邮件客户端发送邮件，如 OUTLOOK,FOXMAIL。

可自定义设置终端 PC 外设端口的开启和关闭，当端口关闭后，端口将不可工作，不可进行信息的传输，如，USB,光盘驱动设备，串口驱动设备，并口驱动设备，蓝牙驱动设备，红外驱动设备等。

4.4.2关键数据加密使用防护

在不影响本地原有明文文档的情况下，无障碍使用密文文档。敏感数据在加密前后对于数据合法使用者无任何差异，不增加用户负担、不改变任何工作流程及使用习惯。文件的保存加密、打开解密完全由后台加解密驱动内核自动完成，对用户而言完全透明、无感知。

可根据文档密级按照组织架构（部门、用户、项目组等）对文件进行密级标识及授权管理，只允许合法授权用户根据分配权限受控使用；非授权用户即使获取到数据也将无法查阅。

对文档设置只读、打印、修改、再次授权、阅读次数及生命周期等权限，授权用户只能按照规定好的权限进行使用，无法通过属性修改、内容复制、副本另存等方式越权使用。

通过内容复制/粘贴、拖拽、副本另存为、截屏/录屏、打印等方式对文档内容进行移植及转储，需要对用户上述操作行为进行安全控制。

4.4.3数据离网交换安全

当数据需要与第三方进行合作编辑，或交由第三方审阅时，为了保证数据在交互中不会被泄露，应采取全面的数据隔离机制来达到安全的效果。

利用沙盒技术与透明加解密相结合，创建一个完全隔离的安全空间，再将需要交换的数据放入此空间，只有拥有密钥和权限的人员才能打开此空间。

认证方式包括：口令、KEY、机器码、口令+机器码。

权限包括：只读、编辑、拷屏、打印、水印、限时、限次、自动销毁。

4.4.4用户上网行为监管

上网行为管理主要解决内部员工上网带来的工作效率低下、带宽滥用、恶意软件感染、内部信息泄漏以及法律合规等问题。

员工通过网络可以查找资料、沟通交流和从事电子商务等，但是相应的多种问题伴随而生，例如：

- 与工作无关的 P2P 和在线视频等应用占用带宽 ；
- 与工作无关的聊天、炒股、游戏、博客和在线购物等活动影响工作效率 ；
- 员工随意在网络上发帖和传输文件导致机密泄露 ；
- 员工上网容易受到病毒、木马和蠕虫等攻击和感染 ；
- 员工通过网络从事一些黄赌毒等违法活动 ；
- 员工浏览和发布不良言论导致企业面临法律风险 。

为解决以上一系列的问题，绿盟科技凭借在应用识别库和网络安全等全方面的多年积累，可实现员工上网行为的管理、带宽的限制和确保员工上网安全等，可以极大提高员工的办公效率和带宽利用率，防止机密数据泄密和员工通过网络从事违法活动。

4.4.5网络数据外发监控

支持多协议的敏感信息识别与监控能力。可以监控包含附件的 SMTP，包含上传下载文件的 HTTP，FTP，还有包含上传文件的 NNTP。还能够监控主流的 IM 协议（QQ,RTX），可正确分辨 HTTP 隧道中的 IM 流量。

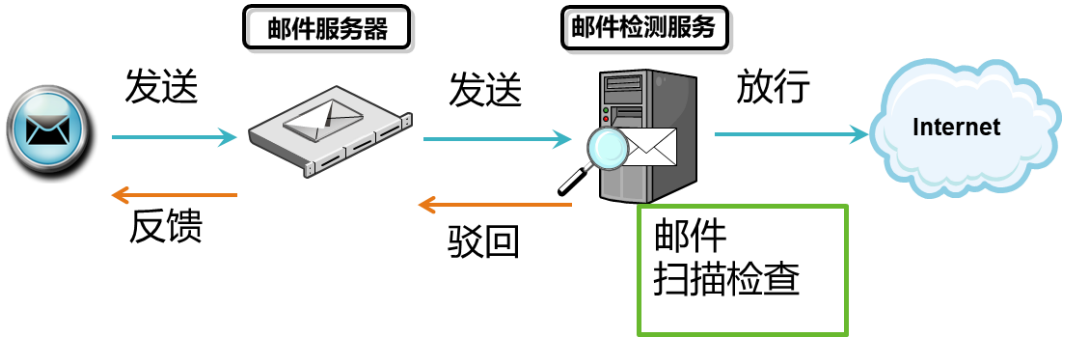
在一个事件产生时，系统可以自动的发送邮件警告给用户、用户经理，IT 管理员。邮件的主题，信体等内容均可以定制。

4.4.6邮件数据安全防护

采用内容识别安全技术，对企业内部对外发送邮件进行实时检测，发现是否包含企业敏感信息，并进行有效的防护。

实现对 POP3、SMTP、IMAP 等主流邮件协议进行分析。检测范围覆盖测内容包括主题、正文、附件等内容。

实现审计、告警、阻断、脱敏、剥离附件、增加收件人、删除收件人、正文替换、邮件审批等多种响应规则，用户可根据管理需要灵活选择。



4.5 数据可视化展现

通过数据溯源和用户实体行为分析来完成可视化的展现，使用户可以更全面更具体的了解数据在全生命周期过程中的状态及风险，为后续防护措施提供有力支撑。

数据溯源由系统创建的四维模型组成，此模型将溯源看成一系列离散的活动集，这些活动发生在整个工作流生命周期中，并由四个维度(时间、空间、层和数据流分布)组成，四维溯源模型通过时间维区分标注链中处于不同活动层中的多个活动，进而通过追踪发生在不同工作流组件中的活动，捕获工作流溯源和支持工作流执行的数据溯源。

用户实体行为分析关联了用户活动和其它实体，如受控或非受控的终端，应用（包括云、移动和其它内部应用）、网络和外部威胁，结合实体的数据本身，通过运用建模、数据分析等手段逐步建立威胁态势感知，使组织能够更好地阻止、检测和应对不断变化的内部威胁。

5 方案关键技术

5.1.1检测与准确性

为了预防数据丢失，无论数据的存储、复制或传输位置在哪里，都必须准确地检测所有类型的机密数据。如果没有准确的检测，数据安全系统就会生成许多误报（将并未违规的消息或文件标识为违规）以及漏报（未将违反策略的消息或文件标识为违规）。误报会大量耗费进行进一步调查和解决明显事故所需的时间和资源。漏报会掩盖安全漏洞，导致数据丢失、潜在财务损失、法律风险并有损组织声誉。

检测技术概述：

为了确保最高的准确性，采用了三种基础检测技术和三种高级检测技术：

- 基础检测技术：

  正则表达式检测（标示符）

  关键字和关键字对检测

- 文档属性检测

基础检测方法采用常规的检测技术进行内容搜索和匹配，比较常见的都是正则表达式和关键字，此两种方法可以对明确的敏感信息内容进行检测；文档属性检测主要是针对文档的类型、文档的大小、文档的名称进行检测，其中文档的类型的检测是基于文件格式化

进行检测，不是简单的基于后缀名检测，对于修改后缀名的场景，文件类型检测可以准确的检测出被检测文件的类型，目前支持 100 多种标准的文件类型，并且可以通过自定义特征，去识别特殊的文件类型格式的文档。

高级检测技术：

- 精确数据比对 (EDM)
- 指纹文档比对 (IDM)
- 向量机分类比对 (SVM)

EDM 用于保护通常为结构化格式的数据，例如客户或员工数据库记录。IDM 和 SVM 用于保护非结构化的数据，例如 Microsoft Word 或 PowerPoint 文档。对于 EDM、IDM、SVM 而言，敏感数据会先由企业标识出来，然后再由 DLP 判别其特征，以进行精准的持续检测。判别特征的流程包括访问和检索文本及数据、予以正规化，并使用不可逆的打乱方式进行保护。

数据检测是以实际的机密内容为基础，而非根据文件本身。因此，此技术不只能检测敏感数据的检索项或衍生项，而且能够标识文件格式与特征信息格式不同的敏感数据。例如，如果已经判别出机密 Microsoft Word 文档的特征，DLP 就能够在相同的内容以 PDF 附件的方式通过电子邮件进行提交时，将其准确检测出来。

精确数据比对：

精确数据比对 (EDM) 可保护客户与员工的数据，以及其他通常存储在数据库中的结构化数据。例如，客户可能会撰写有关使用 EDM 检测的策略，以在消息中查找“名字”、“身份证号”、“银行帐号”或“电话号码”其中任意三项同时出现的情况，并将其映射至客户数据库中的记录。

EDM 允许根据特定数据列中的任何数据栏组合进行检测；也就是在特定记录中检测 M 个字段中的 N 个字段。它能够在“值组”或指定的数据类型集上触发；例如，可接受名字与身份证号这两个字段的组合，但不接受名字与手机号这两个字段的组合。

由于会针对每个数据存储格存储一个单独的打乱号码，因此只有来自单个列的映射数据才能触发正在查找不同数据组合的检测策略。例如，有个 EDM 策略请求“名字 + 身份证号 +手机号”的组合，则“张三”+“13333333333”“110001198107011533” 可触发此策略，但是即使 “李四”也位于同一数据库中，“李四”+“13333333333”“110001198107011533”也不能触发此策略。EDM 也支持相近逻辑以减少可能的误报情形。对于检测期间所处理的自由格式文本而言，单个特征列中所有数据各自的字数均必须在可配置的范围内，方可视为匹配项。例如，依默认，在检测到的电子邮件正文的文本中，“张三”+“13333333333”“110001198107011533”各自的字数必须在选定的范围内，才会出现匹配项。对于含有表式数据 (例如 Excel 电子表格) 的文本而言，单

个特征列中所有数据都必须位于表式文本的同一行上，方可视为匹配项，以减少整体误报情形。

指纹文档比对：

“指纹文档比对”(IDM) 可确保准确检测以文档形式存储的非结构化数据，例如 Microsoft Word 与 PowerPoint 文件、PDF 文档、财务、并购文档，以及其他敏感或专有信息。IDM 会创建文档指纹特征，以检测原始文档的已检索部分、草稿或不同版本的受保护文档。

IDM 首先要进行敏感文件的学习和训练，拿到敏感内容的文档时， IDM 采用语义分析的技术进行分词，然后进行语义分析，提出来需要学习和训练的敏感信息文档的指纹模型，然后利用同样的方法对被测的文档或内容进行指纹抓取，将得到的指纹与训练的指纹进行比对，根据预设的相似度去确认被检测文档是否为敏感信息文档。这种方法可让 IDM 具备极高的准确率与较大的扩展性。

向量机分类比对：

支持向量机（Support Vector Machines）是由 Vapnik 等人于 1995 年提出来的。之后随着统计理论的发展，支持向量机也逐渐受到了各领域研究者的关注，在很短的时间就得到很广泛的应用。支持向量机是建立在统计学习理论的 VC 维理论和结构风险最小化原理基础上的，利用有限的样本所提供的信息对模型的复杂性和学习能力两者进行了寻求最佳的折衷，以获得最好的泛化能力。SVM 的基本思想是把训练数据非线性的映射到一个更高维的特征空间（Hilbert 空间）中，在这个高维的特征空间中寻找到一个超平面使得正例和反例两者间的隔离边缘被最大化。SVM 的出现有效的解决了传统的神经网络结果选择问题、局部极小值、过拟合等问题。并且在小样本、非线性、数据高维等机器学习问题中表现出很多令人注目的性质，被广泛地应用在模式识别，数据挖掘等领域。

SVM 比对算法适合那些具有微妙的特征或很难描述的数据，如财务报告和源代码等。

使用过程中，先将文档按照内容细分化分类，每一类文档集合有属于本类的意义，经过 SVM 比对，确定被检测的文档属于哪一类，并取得此类文档的权限和策略。

同时，针对 SVM 的特点，可以进行终端或服务器上的文档按照分类含义进行分类数据发现。

IDM 和 SVM 的比对区别是，IDM 将待检测文件的指纹和训练模型中的每一个文件进行指纹比对；而 SVM 是将待检测文件向量化，并归属到某一类训练集所建立的向量空间。

5.1.2 多语言和语义的检测支持

提供多种语言的检测支持，简体中文、繁体中文、日文、韩文、英文。同时也提供多种语言的语义识别。

5.1.3文件级智能动态加解密技术

一种文件级过滤驱动编程技术，其发展历经三个阶段：单缓存过滤驱动技术、双缓存过滤驱动技术和虚拟文件系统技术（LayerFSD）。目前商业市场上大多数内核级加密厂商均采用单缓存过滤驱动技术，少量厂商已发展到双缓存过滤驱动技术，而发展到虚拟文件系统技术（LayerFSD）并实现产品化的厂商则屈指可数，绿盟科技公司早在2009年就实现LayerFSD技术。

文件过滤驱动技术，通过实时拦截文件系统的读/写请求，对文件进行动态跟踪和透明加/解密处理。其主要优点：文件加/解密动态、透明，不改变使用者的操作习惯；性能影响小，系统运行效率高；不改变原始文件的格式和状态，同时，部署和内部使用非常方便。

显著特征为：加密强制性、使用透明性、保密彻底性、应用无关性、灵活拓展性。

5.1.4设备过滤驱动技术

一种设备过滤驱动编程技术，可实现对终端任意设备(USB端口、打印机、光驱、软驱、红外、蓝牙以及网卡等)的安全保护及控制。绿盟科技公司于2009年成功研发并完善该技术。

6 方案价值

1） 满足合规要求

现如今，国家对数据安全已经出台了多项法规，通过本方案的实施，可以对法规中提到的鉴别信息数据、重要个人信息、重要业务数据做到针对性的监控与保护，使企业在发现数据风险前及时做出响应，避免因数据丢失造成的危害与损失。

2） 权限划定清晰

责权不清一直都是最根本的问题，通过本方案的实施，将数据合理的进行级别划分，再结合管理与业务的需要对数据的访问、使用，进行清晰的权限管控，做到权责分离，事后还可以通过审计结果明确事故责任方，避免了责任不清出现的推诿扯皮。

3） 数据生命周期全面掌控

掌握数据的全生命周期是对数据风险的提前预知，利用本方案对数据的生命周期中各个环节做监控，掌握数据的动态，了解数据的流向，提前对可能发生的数据泄露风险进行预警，保障数据在安全的可控范围内流转、使用与存储。

4） 降低数据泄露风险

通过对数据的扫描与跟踪，利用内容识别、UEBA、机器学习等技术，及时发现数据所承载的系统、业务、网络、终端中的安全威胁，提前做好防范措施，让泄密风险看得见、使数据泄漏防得住。

5） 提高数据使用者的安全意识

绿盟数据安全解决方案的应用，让数据使用者了解数据的重要程度，规范数据使用者的操作行为，从潜意识里指导与帮助人们正确使用资源，合理利用资源，保护数据的安全。