FUJIAN TELECOMMUNICATIONS TECHNOLOGY

福建通信科技

二〇二二年
论文集

第三届“华安星杯”
网络与数据安全优秀解决方案

福建省通信学会
福建省互联网协会
福建省信息通信行业协会
福建省邮电规划设计院有限公司

闽内资准字K第111号
内部资料　免费交流

序　言

为深入贯彻落实习近平总书记关于网络强国的重要思想，围绕党的十九大以来网络空间安全领域的科研成果，分享有关网络与数据安全研究成果和应用经验，广泛探讨网络空间安全所面临的风险挑战问题，促进技术创新，推动网络与数据安全产业发展，在福建省科学技术协会、中国通信学会、福建省通信管理局和福建省工业和信息化厅的指导下，由福建省通信学会、福建省互联网协会、福建省网络与信息安全产业发展促进会、福建省信息协会、福建省计算机学会、福建省高校教育信息化学会、福建省卫生信息协会、福建省互联网信息交流协会、福建省互联网金融协会共同主办，福建中信网安信息科技有限公司、福州市晋安区委人才工作领导小组办公室、福州市晋安区科学技术协会共同承办，于 7 月 1 日起面向福建省内企事业单位、省内高校（含高职）公开征集福建省 2022 年东南科技论坛——数据安全与数字经济产业融合发展论坛第三届“华安星杯” 网络与数据安全优秀解决方案。

活动期间，共收到来自省内高校、运营商、网络安全企业、科研机构、医疗卫生等单位的解决方案作品 25 篇，内容涵盖数据安全、网络安全两大板块，涉及 5G、人工智能、北斗网格码、零信任、微隔离、隐私计算等技术领域。由福建省互联网网络与信息安全专家组成的作品评审专家组遵循“公平、公正、公开”的原则，从解决的方案的立意创新性、撰写的规范性、研究的先进性、数据和结论的可信度以及推广示范应用价值等角度展开详细评审。经过深入研讨，最终综合评议出 12 篇优秀解决方案，并与其他 13 篇解决方案作为《福建通信科技》（闽内资准字 K 第 111 号）增刊发行。

在此特别感谢积极参与此次征集活动的投稿作者，同时也对评审专家、《福建通信科技》编委会、指导单位、主办单位、承办单位等工作人员的辛勤付出表示衷心的感谢！

编 者

2022 年 9 月 13 日

目 录

数字证据链的多维关联度分析方法

邓文涛[1] 陈洪[2] 张明辉[2] 林睫菲[2] 刘延华[1]
（1.福州大学 计算机与大数据学院，福建 福州 350108；
2.国网信通亿力科技有限责任公司，福建 福州 350003）

摘 要：受数字证据的海量性、关联复杂性和动态可变性等特点影响，现有的犯罪取证工作存在着数字证据分散，取证分析效率低等问题，降低了数字证据的实际证明力，影响数字证据在实际审判中的采纳程度。本文提出一种基于多维关联度的数字证据分析方法。首先，设计一种数字证据标准化表示方法，将数字事件和其之间的关联关系进行规范化描述。然后通过对数字事件关联关系的分析，提出了多维度关联度计算方法以及基本证据环的构造分析。实验结果表明，所提出方法构造的数据证据链对于提升数字证据的证明力具有一定的应用意义。

关键词：数字取证；数字证据链；证据环；关联度

0 引言

数字取证是一种收集和分析数字证据的法律程序，目的是将嫌疑人和数字犯罪活动联系起来[1-2]，已成为数字犯罪调查的重要部分。数字取证的核心就是反复分析证据事件之间关联性，重复、回溯地建立更多数字证据链[3]的过程。

数字证据链描述了数字证据之间的相互作用及逻辑关联性。在案件调查中，可能需要提供不同结构的证据链，它们的证明力也是不同的[4]。因此，如何构造有效的多样性的数字证据链，成为数字证据研究的重要内容。

围绕数字证据链的综合分析，本文研究了数字证据链的多维关联分析方法，实现了数字证据的有效构造。主要研究工作与贡献如下：

（1）提出了数字证据的标准化表示方法，解决了数字证据链中主客体、数字事件以及多种关联关系的规范化表达，为数字证据关联度的表示与计算提供了重要支持。

（2）研究了数字事件的多维综合关联度的计算方法，提出基于时间间隔的时间关联度、基于Jaccard 系数的主客体关联度、基于 LDA 算法的内容关联度等多种关联度的计算方法，实现了数字事件之间关联度的自动计算。

1 相关工作

1.1 抽象数字事件的重构方法

Somayeh Soltani 等[5]研究通过使用文件系统元数据中的应用程序来重建高级事件，并提出了一个事件重构框架，用于确定哪些应用程序已在受损系统上运行。伏晓等[6]针对计算机入侵取证中的入侵事件场景重构技术进行了研究。Song S[7]针对 Android 网络钓鱼攻击问题，提出了一个应用数字取证工具，该工具可以成功地在安卓设备上重构攻击的场景。此外，还有一些研究工作关注了数字证据链的可视化技术[8]，在证据链的呈现方面取得了不错的效果。

1.2 基于时间线的证据链研究

时间线分析方法是研究证据链的基本方法，Yoan Chabot 等[9]提出了事件的时间线构建模型，设计了事件重构中关联度计算方法，实现了事件之间的时间、主客体、规则等多种关联度的量化计算，并以形式化和实例两种方式验证了方法的有效性，对于证据链的关联性分析具有很好的参考价值；Hargreaves 等[10]研究了证据链中基本事件的融合方

法，并提出了基于模式匹配的数字事件自动重建框架，为证据链的超级时间线分析做了一些有益探索。为了分析Windows系统中的数字事件的发展态势，Yuandong Zhu[11]提出了基于 Windows 还原点状态比较的数字事件的时间线分析方法，使得数字事件的改变更容易被发现。

2 数字证据的规范化表示

一个数字证据是由一个或多个数字事件及其相关主体、客体等关联而成的。而一个数字事件是表示是若干主体对若干客体实施的某些操作。

2.1 主客体的表示

一个主体 s_i，包括以下几个部分组成：

（1）s_id：识别主体的唯一性编号。

（2）s_name：主体的名称，如 Firefox、Foxmail、Windows 等。

（3）s_user：运行主体的系统用户名称或社会用户本人；

（4）s_image：主体如果是进程，该部分表示其对应的硬盘映像；如果主体是社会用户则为其本人。

（5）s_host：主体如果是进程，该部分表示进程所在的设备或系统名称；如果主体是社会用户则为本人。

一个数字事件可能涉及多个主体，此时事件的主体集合可以描述为：$S=\{s_1,s_2,...,s_n\}$。

一个客体 o_j，包括以下几个部分组成：

（1）o_id：识别客体的唯一性编号。

（2）o_name：客体的名称，如 file1、pic1 等。

（3）o_source：客体的来源，如一张照片采集于某主机或某网站等。

（4）o_tc：客体被创建的时间或最早被发现的时间。

（5）o_tu：客体最后一次被更新（修改）的时间。

（6）o_tr：客体最后一次被读取的时间。

（7）o_device：客体的原始生成来源，如一张照片可能由一部相机拍摄产生。

（8）o_owner：客体的属主，即客体的拥有者，可能是一个用户或机构等。

（9）o_user：最后访问客体的用户名称。

一个数字事件所涉及的客体集合描述为 $O=\{o_1,o_2,...,o_m\}$。

2.2 数字事件的表示

事件集合表示为 $E=\{e_1,e_2,...,e_l\}$，一个事件 e_k 主要包括以下几个部分：

（1）e_id：识别事件的唯一性编号。

（2）tstart：事件开始或最早出现的日期时间。

（3）tend：事件结束或最迟出现的日期时间。

（4）e_local：事件发生的地点，如某一主机、一个地方等。

（5）S_e：事件所涉及到的所有主体的集合。$S_e=\{s\in S\mid e\in E,s\ \ r_s\ \ e\}$，$r_s$ 表示主体与事件之间的关联关系，r_s 的定义见下文。

（6）O_e：事件所涉及到的所有客体的集合。$O_e=\{o\in O\mid e\in E,o\ \ r_o\ \ e\}$，$r_o$ 表示客体与事件之间的关联关系，r_o 的定义见下文。

（7）E_e：与该事件相关的所有其他事件的集合。$E_e=\{e\in E\mid x\in E,x\ \ r_e\ \ e\}$，$r_e$ 表示事件之间的关联关系，r_e 的定义与计算方法见下文。

（8）e_info：描述事件的一段文字或若干个关键词，可用于基于主题和内容的检索、挖掘或相似度计算等数字证据分析。

2.3 关联关系的表示

关联关系是数字证据链中的关键信息，一个证据链的证明力大小往往由其包含的关联关系所决定。

（1）数字事件与主体之间的关联关系 r_s

数字事件和主体之间存在直接参与和间接影响的关系，令 r_s={参与,影响}。

（2）数字事件与客体之间的关联关系 r_o

客体最常见的对象形式为数据或文件，通过借鉴文件的访问方式，令 r_o ={创建,修改,删除,读取,终止}。

（3）事件与事件之间的关联关系 r_e

与 r_s 和 r_o 相比，数字事件之间的关联关系 r_e 更为复杂，包括基于属性域的关联、基于时间关系的关联、基于因果关系的关联、基于事件内容的关联等。

那么，数字事件之间的关联关系定义为：r_e ={主体关联度,客体关联度,时间关联度,内容关联度}。

3 数字事件的关联度计算

3.1 时间关联度的计算

在取证中，一些数字证据可能只有一个创建时间或访问时间，这增加了数字事件的结束时间或发生周期等属性的确定难度。采用艾伦代数（Allen algebra）表示法[12]来表示数字事件之间的时间间隔。

两个事件间的时间间隔计算如下：

$$Ts = y_{t_{start}} - x_{t_{end}} \quad (1)$$

给定一个时间阈值 τ，当 Ts 超过 τ，则将两个事件的时间关联度设置为 0。τ 的大小视数字证据事件的类型而定，如调查 Web 浏览事件和 Email 行为时，τ 的取值是不同的。

当数字事件之间的时间间隔小于 τ 时，用关联度函数 $r_{et}(x,y)$ 来对两个事件之间的时间间隔进行量化度量，计算结果称为超时间线的时间关联度。

假设事件 x 发生在事件 y 之前，则 $r_{et}(x,y)$ 的计算如下：

$$r_{et}(x,y) = start(x,y) + equal(x,y) + meet(x,y) + overlap(x,y) + during(x,y) + finish(x,y) + before(x,y) \quad (2)$$

在公式(2)中，除了 $before(x,y)$ 外，其他函数均为二值函数，当满足如文献[12]给出的对应约束条件时，函数的值为 1，否则值为 0。

当 Ts 的取值范围是 $(0, \tau)$ 时，函数 $before(x,y)$ 的计算如公式（3）所示：

$$before(x,y) = 1 - (\frac{Ts}{\tau})^2 \quad (3)$$

从数字取证分析的一般规则来看，当两个数字事件同时发生或发生周期相同，即满足 $equal(x,y)$ 或 $start(x,y)$ 时，则认定它们之间的时间关联性更强。

为此，我们设定一个大于 1 的常数 β，来增强这两个函数在时间关联度计算中的权重，此时函数 $r_{et}(x,y)$ 的计算公式就更新为：

$$r_{et}(x,y) = \beta \times start(x,y) + \beta \times equal(x,y) + meet(x,y) + overlap(x,y) + during(x,y) + finish(x,y) + before(x,y) \quad (4)$$

3.2 基于 Jaccard 系数的主客体关联度计算

数字事件由多个属性域组成，当两个事件的属性域具有同一数值或指向同一对象时，则采用 Jaccard 系数来计算事件之间的主客体关联度。

设J(A, B)表示 Jaccard 系数，其定义如下：

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

其中，$|A \cap B|$表示两个集合中的交集中的元素的数目，$|A \cup B|$则表示两个集合的并集中元素的数目。

（1）基于同一主体的关联性度计算

每个事件的 S_e 都包含若干个主体，两个事件之间拥有共同主体数目所占的比值来作为它们的关联度。

即事件 x 和 y 的主体关联度计算如下：

$$r_{es}(x,y) = \frac{|S_x \cap S_y|}{\max(|S_x|,|S_y|)} \quad (6)$$

其中，$|S_x|$ 表示数字事件 x 包含的所有主体的数目，$|S_y|$ 表示数字事件 y 包含的所有主体数目，$|S_x \cap S_y|$则表示数字事件 x 和数字事件 y 共同包含的主体的数目。

（2）基于同一客体的关联度计算

每个数字事件的 O_e 都包含若干个客体，这里将用两个数字事件之间拥有的共同客体数目所占的比值来计算它们之间的关联度。

事件 x 和 y 之间的客体关联度计算公式如下：

$$r_{eo}(x,y) = \frac{|o_x \cap o_y|}{\max(|o_x|,|o_y|)} \quad (7)$$

其中，$|O_x|$ 表示事件 x 包含所有客体的数目，$|O_y|$ 表示事件 y 包含的所有客体数目，$|O_x \cap O_y|$则表示事件 x 和事件 y 共同包含的客体数目。

3.3 基于 LDA 的事件内容关联度计算

事件的 *e_info* 是由文本或若干个主题词来描述的，通过 LDA 主题模型计算事件内容的相似度，来量化两个数字事件之间的内容关联度，计算步骤如下：

Step1：获取两个事件的 *e_info* 的内容；

Step2：利用 LDA 主题模型，对两个数字事件的 *e_info* 内容进行主题词提取，并将获得的主题词及其分布概率转化为主题词权值对，如表 1 所示：

表 1 两个事件的内容主题词权值对

主题词集合 T	主题词权值对	
	在事件 x 中的权值	在事件 y 中的权值
T1	ω_{x1}	ω_{y1}
T2	ω_{x2}	ω_{y2}
……	……	……
Tn	ω_{xn}	ω_{yn}

其中，主题词集合 $T=\{T1,T2,...,Tn\}$，由两个事件 *e_info* 中主题词组合生成。对每个主题词 T_i，在事件 x 的 *e_info* 中的概率为 ω_{xi}，在事件 y 的 *e_info* 中的概率为 ω_{yi}。

Step3：由表中的权值对数据，计算事件内容的相似度 $r_{ek}(x,y)$，公式如下：

$$r_{ek}(x,y)=\frac{\sum_{i=1}^{n}(1-(\omega_{xi}-\omega_{yi})^2)}{n} \quad (8)$$

至此，设计了数字事件之间的多个关联度计算方法，包括 $r_{es}(x,y)$、$r_{eo}(x,y)$、$r_{et}(x,y)$ 和 $r_{ek}(x,y)$ 等，从不同角度呈现了数字证据链中的数字事件的联结关系。

为了更加综合地分析两个数字事件之间的关联关系，定义一个**综合相似度函数** $r(x,y)$，将它们之间的不同关联度进行加权求和，公式如下：

$$r(x,y)=\omega_s \times r_{es}(x,y)+\omega_o \times r_{eo}(x,y)+\omega_t \times r_{et}(x,y)+\times \omega_k \times r_{ek}(x,y) \quad (9)$$

式中，不同关联度对应的权重系数 ω_s、ω_o、ω_t 和 ω_k 可由领域专家设定。

4 实验结果与分析

4.1 实验数据

为了验证所提出的证据关联度计算方法和数字证据链构造方法，本节以 DFRWS 2013 年会挑战赛（DFRWS The 2013 Data Sniffing Challenge）发布的 Web 浏览数据(即 txt-04:975KB)[13]和自行采集的 Web 浏览数据集开展实验。

4.2 基于时间和主客体关联度的证据链构造

根据 $r_{et}(x,y)$、$r_{es}(x,y)$、$r_{eo}(x,y)$ 的计算公式，计算数字事件之间的对应关联度，并用有向图形式来表示实验结果，每个节点代表一个浏览事件，有向边则表示事件之间的时间先后关系，事件之间的关联度大小作为边的值。

部分证据链如图 2、图 3 所示：

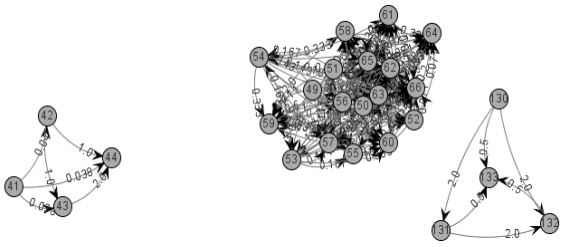


图 2 基于时间关联度的证据链

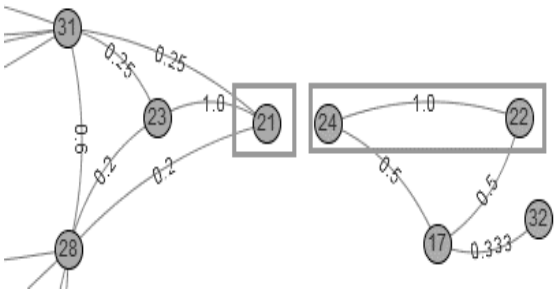


图 3 基于客体关联度的证据链

4.3 基于内容关联度的证据链构造

本实验使用自行采集的 Web 浏览数据集进行实验。

为了便于分析数字事件之间的关系，设置一个内容关联性约束的阈值，减少数字事件的数量。图 4 所示的是内容关联度阈值设置为 0.9 的结果图。

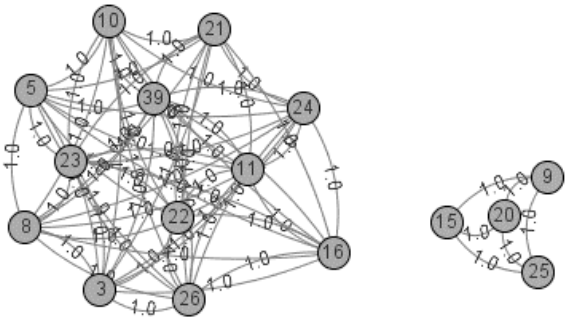


图 4 内容关联度大于 0.9 的数字证据链

在实际取证工作中，取证专家可以根据需要设置具体的内容关联度阈值，查看哪些数字事件具有更好的内容相关性。

结束语

本文围绕数字证据链的自动构造与分析问题，在对数字证据规范化表示的基础上，提出了基于多维关联度计算的数字证据链构造方法。通过对数字事件关联关系的分析，对不同维度关联关系的计算进行了具体的定义。与已有研究工作相比，提出的数字证据关联度分析丰富更加丰富，且能够进行自动化计算，对于实现数字证据链的自动化分析与构造具有积极作用。

参考文献：

[1] Lin X, Chen T, Zhu T, et al. Automated forensic analysis of mobile applications on Android devices[J]. Digital Investigation, 2018, 26: S59-S66.

[2] Lone A H, Mir R N. Forensic-chain: Ethereum blockchain based digital forensics chain of custody[J]. Sci. Pract. Cyber Secur. J, 2018, 1: 21-27.

[3]马国富,王子贤,王揆鹏.基于证据链的电子司法鉴定模型[J].河北大学学报(自然科学版),2013,33(3):317-323.

[4]刘文彦,霍树民,陈扬,等.网络攻击链模型分析及研究[J].通信学报,2018,(z2).88-94.

[5]Somayeh Soltani and Seyed Amin Hosseini Seno and Hadi Sadoghi Yazdi.Event reconstruction using temporal pattern of file system modification Security,2019,13(3):201-212.

[6]季雨辰,伏晓,石进,骆斌,赵志宏.计算机入侵取证中的入侵事件重构技术研究[J].计算机工程,2014,40(1): 315-321.

[7] Song S, Liu X, Fu X, et al. Visible Forensic Investigation for Android Applications by Using Attack Scenario Reconstruction[C]//2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021: 1-6.

[8] David Gugelmann,Fabian Gasser,Bernhard Agerc,et al.Hviz: HTTP(S) traffic aggregation and visualization for network forensics. Digital Investigation, 2015,12S:S1-S11.

[9] Yoan Chabot, Aurelie Bertaux, Christophe Nicolle, et al.A complete formalized knowledge representation model for advanced digital forensics timeline analysis. Digital Investigation, 2014,11S:S95-S105.

[10] Christopher Hargreaves, Jonathan Patterson. An automated timeline reconstruction approach for digital forensic investigations.Digital Investigation,2012,9:S69-S79.

[11]Yuandong Zhu,Joshua James,Pavel Gladyshev.A comparative methodology for the reconstruction of digital events using windows restore points. Digital Investigation,2009,6: 8-15.

[12] JAMES F.ALLEN.Maintaining knowledge about temporal intervals.Communications of the ACM,1983,26(11):832-843.

[13]DFRWSS.DFRWSs-13-challenge-ests.zip [Online]. https://github.com/dfrws/dfrws2012-2013-challenge.

基金项目：国家自然科学基金(62072109, U1804263)；福建省自然科学基金(2021J01625,2021J01616)。

通信简介：刘延华，博士，副教授，主要研究方向为网络空间安全、智能计算及应用等。

数据开放共享安全解决方案

高 垠　李剑飞　陈惠源
（福建极推科技有限公司，福建 福州 350001）

摘　要：为有效挖掘数据价值的同时保证数据的安全性，本文提出构建数据开放实验室，利用中心化与弱中心化两种模式，以及联邦学习等隐私计算技术，完成数据在提供方本地使用，完成加密联合建模的解决方案。不仅提高数据共享和业务协同能力，同时确保数据不出域、保障数据隐私权，旨在解决数据在受控环境下的价值挖掘问题，在政府监管下，拉通数据供需，规避数据权属界定问题，既实现了数据要素的价值挖掘，又充分地保障了数据安全。

关键字：数据共享；业务协同；数据安全

1 背景

第四次工业革命以数字化、智能化、网络化为核心，而数据资源作为重要生产要素，蕴藏了巨大的价值，被认为是 21 世纪的“黄金”“石油”。2019 年，党的十九届四中全会首次将数据与土地、劳动力、资本、技术、并列作为重要的生产要素。2020 年，《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》和《中共中央国务院关于新时代加快完善社会主义市场经济体制的意见》均强调要培育和发展数据要素市场。2021 年，“十四五”规划《纲要》进一步明确提出，要建立数据资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范，推动数据资源开发利用。

数据要素的特殊属性，要求加强数据资源的开放共享。数据越多价值越大，越分享价值越大，越不同价值越大，越跨行业、区域、国界价值越大。因此，实施数据开放共享，优化治理基础数据库，不断完善数据权属界定、开放共享、交易流通等标准和措施，促使数据资产重复使用、多人共同使用、永久使用，加快推动各区域、部门间数据共享交换，显得十分必要。

但由于政府、企业各部门间不同系统和业务的闭塞性和阻隔性，数据信息共享困难、数据孤岛问题严重；而通过分布式建模计算整合数据源的方式，在数据的传输和处理方面效率低下；另外随着人工智能技术进一步发展和应用，数据隐私问题日益突出，“数据泄露、数据贩卖”等数据安全事件的频繁发生也给个人、企业、社会带来了巨大影响。就目前而言涉及到数据的隐私保护、数据归属权等难题，在兼顾数据安全和隐私的条件下想要实现数据融通共享，仍处于“不敢”、“不能”、“不愿”状态，在数据合作和融通过程中，广泛面临以下问题：

（1）数据缺乏安全管控手段

数据融通后，数据拥有者失去了数据的控制权，无法对协同出去的数据的使用进行干预和监控，难以对数据使用的行为进行管控审计，如果不能对数据确权，明确数据的产生者、使用者、管理者及受益者，将无法很好实现数据的精准授权，严重阻碍数据的流通以及价值变现。

（2）数据安全保护不足

数据在融通过程中，缺乏安全的计算处理环境，数据在处理过程中可能被非法窃取，或者在计算后，计算结果仍然可能会被泄露，导致数据的拥有或管理方不放心让自己的数据进入公共域。

（3）数据隐私防护不足

由于被共享数据的知情和授权不足，分享的数据很多是明细的“裸数据”，具有极大的安全隐患，数据的拥有者出于隐私考虑，不愿意共享数据，因此如何做好隐私防护，实现数据共享与合作“知情

权”、“最小化利用”成为数据合作价值挖掘的必要前提。

（4）数据融通合作基础设施不完善

首先，各方缺乏统一的数据标准，数据的定义、口径各有不同；其次，缺乏高效便捷的数据合作基础设施，各机构的数据库、数据加工引擎、算法、可视化工具等各不统一，在合作过程中需要投入巨大资源进行系统改造、对接，造成数据合作困难、合作周期长。

2 解决方案

基于固定安全边界保障数据不出域的原则，构建数据开放实验室，打造集数据资源、算力、算法、办公场所等条件可信安全的数据分析及价值挖掘的工具平台，旨在解决在数据安全的情况下进行数据的受控开放以及数据价值挖掘问题。（架构图如图 1-1）

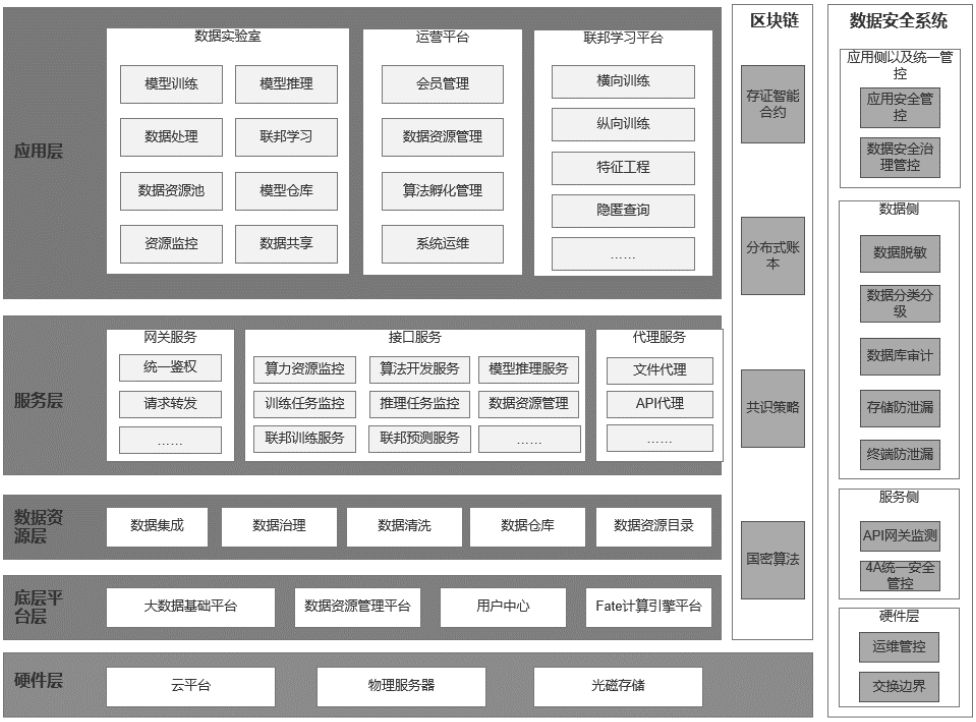


图 1-1 平台架构图

（一）平台架构说明

（1）硬件层

硬件层包括云平台、物理服务器、自主可控的光磁存储介质，为平台提供数据计算、数据存储等能力；同时依托于kubernetes容器化管理平台，实现硬件资源容器编排管理。

（2）底层平台层

底层平台层整合大数据基础平台、数据资源管理平台、用户中心、fate计算引擎平台，打通各大平台或软件模块生态流程。大数据基础平台不仅能提供分布式计算、存储能力，同时也可以提供丰富的技术组件；数据资源管理平台统一纳管结构化数据，为模型训练、模型推理、数据处理、数据共享等应用提供数据源；用户中心提供用户权限管理的能力；fate计算引擎平台能够针对外部不可出域数据，提供联邦学习计算能力。

（3）数据资源层

利用底层平台层提供的工具，在数据资源层完成数据集成、数据治理、数据清洗，进而建设数据仓库，最终形成数据资源目录，实现数据全域治理和全生命周期管理。

（4）服务层

服务层可提供网关服务、接口服务、代理服务这三大服务。网关服务提供统一鉴权、请求转发等

功能;接口服务提供算法资源监控、算法开发服务、联邦训练服务、联邦预测服务等功能；代理服务提供文件代理、API代理等功能。

（5）应用层

应用层具有数据实验室、联邦学习平台等模块，可以针对不同的应用场景提供不同的服务。数据实验室是基于固定安全边界提供数据资源、算力、办公场所等条件的支持易数工场受托服务业务及自身算法孵化的封闭的数据分析及数据挖掘工具；联邦学习平台针对不出域的数据,通过联邦计算的方式，实现数据融通，实现数据可用不可见的安全流通解决方案；运营平台可提供监控系统运维、展示数据资源目录、管控数据资源的开放权限、算法孵化管理等功能。

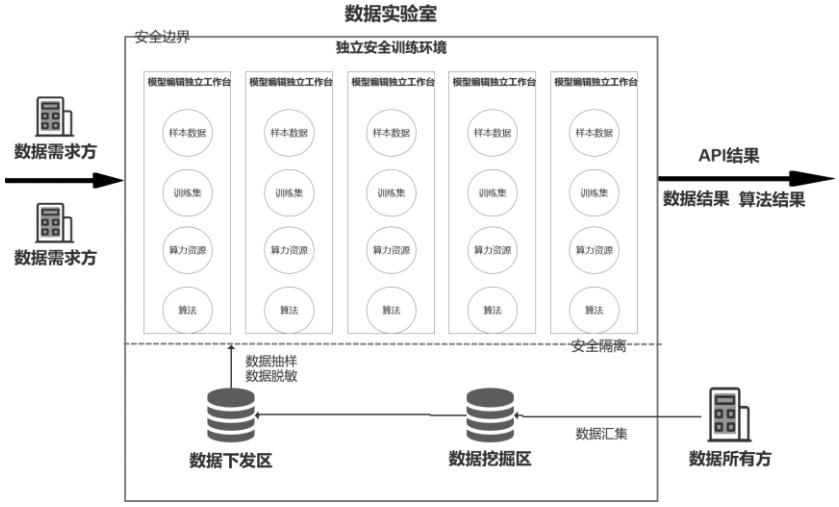
同时为增强数据的可信度和提升审计监管能力，平台在各环节依托区块链存证能力,实现数据审核、应用的全方位监管；为保障数据安全，平台打造完整安全体系：针对硬件及运维侧，有运维管控及安全交换边界；在数据侧，包括数据库审计、脱敏、终端防泄漏、网络防泄漏能力；面向服务侧，则通过API网关监测、4A统一安全管控保障API安全及权限安全；应用侧则依托应用安全管控对应用安全进行保障。在以上工具基础上构建数据安全治理管控平台形成数据全生命周期安全保障，面向用户提供数据安全可视化及管控能力。

（二）核心能力

其中以“中心化算法孵化”及“弱中心化联邦学习”两种模式为核心，以数据及算力资源统一配置管理为根本,以数据处理、数据分析、模型训练、模型预测等算子研发为基础，以不断完善“算法仓库”为重点，对接自研区块链形成数据生命周期全流程监管，从而有效解决数据融通过程中的数据价值安全挖掘的核心问题。

（1）中心化算法孵化模式

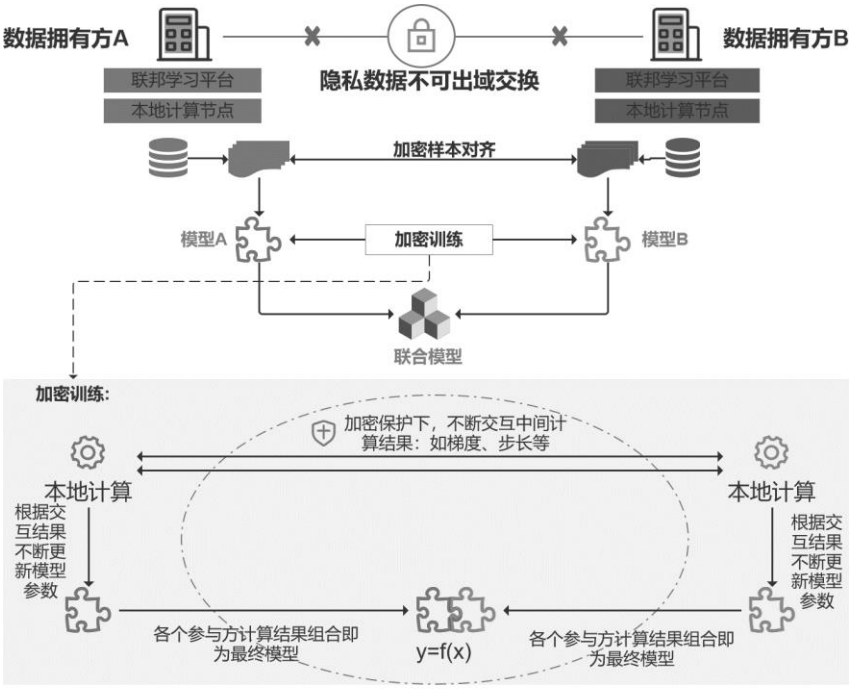
通过封闭的网络环境和固定的物理空间,结合统一的数据资源池、丰富的算力与算法资源，满足用户的数据需求，实现数据的受控开放，并通过区块链技术,实现了数据协同开发过程的全流程存证记录。当数据所有方同意数据使用权释放至平台时可采用该模式进行数据挖掘,通过搭建网络隔离与物理空间隔离的环境结合全方位的数据安全工具充分保障数据安全,通过打造数据资源池与丰富的算力环境,为每位数据需求方提供独立安全的训练环境，数据需求者在平台中进行训练与分析后，将结果通过 API 方式进行交互，原始数据无法获取，训练结束后该需求方的独立工作台包含申请数据、运行中间结果、算力等自动销毁。如下图：



（2）弱中心化联邦学习模式

针对不出域的数据，通过联邦计算的方式，实现数据融通环节通过数据可用不可见技术提供数据可用不可见的安全流通解决方案，释放数据价值，实现产业间高效协同，助力政企数据、企企数据价值的共享与协作。当数据所有方与数据需求方的数据均不可出域，且需要进行联合建模时，可通过平台提供的联邦学习方式，完成数据在拥有方本地使

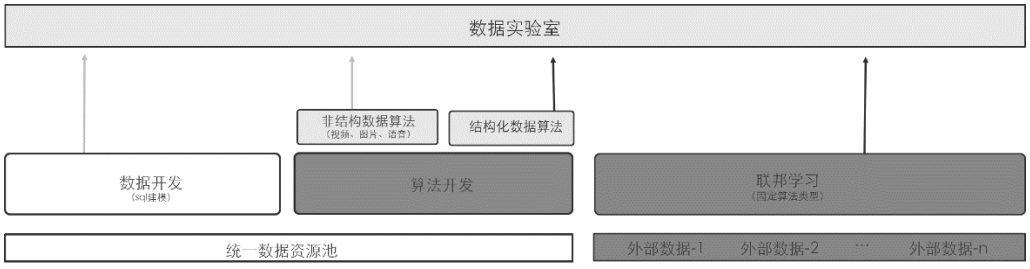
用，完成加密联合建模，通过在数据拥有方部署本地计算节点，通过加密样本对齐、加密训练等，数据拥有方不交换原始数据，仅在加密保护下交互中间计算结果，通过联合建模方式获取最终模型，不仅提高数据共享和业务协同能力，同时确保数据不出域、保障数据隐私权。如下图：



（三）方案优势

（1）低代码一站式数据开发工具

数据开放实验室针对作为封闭的数据价值挖掘工具平台，提供一站式的数据开发工具。支持针对结构化数据的数据分析建模、机器学习算法开发，针对非结构化的（视频、文本、图片）等深度学习的算法开发，同时，针对不出域的数据可以通过联邦计算的方式实现“数据可用不可见”。



数据开放实验室的算法训练工具基于kubernetes,kubeflow 和 docker 虚拟化技术实现训练环境的自动部署和分布式训练环境，并提供了包括数据下载，在线训练，在线验证，算法镜像封装等一站式工具。

数据开放实验室提供 Jupyter 编辑器支持开发者进行算法编辑,还内置许多联邦学习算子与算法，并提供托拉拽面板进行联邦学习训练,通过一站式、图形化算法训练工具大大降低用户的操作难度。

（2）全方位安全管控

在数据开放实验室内部系统及底层平台符合“等保三级”的前提下，数据开放实验室还引入了以下安全技术手段：

a. 网络安全体系优化

在本地政务外网基础上，划分安全域，包括数据汇聚区、数据治理区、数据安全计算区、数据运营区，安全区间配备网络安全隔离设备，如VFW,VIPS,VWAF 等，确保区间数据流、指令流严

格受控。

b. 数据全生命周期安全保障

针对数据接入、存储、传输、计算、交换、销毁流程，配备数据安全分类分级、终端数据防泄漏系统、数据存储防泄漏、网络异常行为阻断、数据库防火墙、数据库脱敏加密等系统，保障数据全生命周期安全。

c. 数据权限控制

以身份为中心，通过帐号管理、认证管理、集中授权、综合审计等软件子系统，实现应用系统及数据资源的最小实体级授权，统一的访问入口维护，运营、运维、技术及数据需求方各类用户角色的操作及生命周期进行管理。

d. 数据安全计算

针对入驻的产业用户和生态服务商，通过固定网络环境、固定操作空间的数据开放实验室进行封闭的数据开发挖掘，借助联邦学习、多方安全计算等核心计算，实现“数据可用不可见”，流通的只是加工处理过的数据结果而非原始的明细数据。

（3）隐私计算安全加持

数据开放实验室通过提供联邦学习实现各参与方无需共享或交换各自敏感数据的情况下进行联合建模，支持横向联邦学习、纵向联邦学习常见模式。内置 LR、GBDT、DNN 等机器学习算法，通过联邦学习进行模型碰撞实现数据隔离：联邦学习的整套机制在合作过程中，数据不会传递到外部，通过同态加密、隐匿计算等为数据实验室不出域的价值碰撞提供了安全加持。

（4）一体化资源配置

数据开放实验室通过前台、后台的方式实现数据资源、算力资源的统一管控、统一配置。产业用户或者生态服务商通过数据资源、算力资源申请的方式进行任务是申请，数据开放实验室平台运营方可以对其任务进行审核并通过工具进行资源配置，实现了数据资源和算力资源的整体运营。

3 应用经验

（1）“数据沙箱”工具助力某市产融平台风控服务

需求背景：

- 响应国家扶贫政策号召，某市建设“三农”产融平台解决农民“贷款难”问题
- 三农业务涉及农宅、农地、农机、种植大户、养殖大户等数据，数据来源农业局、林业局，需要协调数据出域问题，另外也需进行数据治理工作。

解决方案：

当地大数据局协调农业局、林业局数据，以“数据沙箱”形式将数据资源、算力资源、数据开发工具提供给数据开发者，数据开发者完成模型开发后提交政府审批，审批通过后以API形式提供数据服务至数据需求方（产融平台运营方），后续API调用过程、授权信息通过区块链留痕确保后续可追溯，从而实现“数据开发者”、“数据需求方”在不持有数据的前提下使用数据。

（2）“联邦学习”工具助力某助贷公司实现精准营销

需求背景：

- 某市助贷公司需使用不动产、公积金等数据帮助银行对存量客户进行二次营销；
- 助贷公司可获取用户手机号、身份证号等信息；
- 政府侧担忧不动产、公积金等涉及公民财产数据开放引发数据安全风险及隐私泄露问题。

解决方案：

在大数据局侧和助贷公司使用数据实验室中联邦学习平台，数据分别上传至各自域内节点，助贷公司通过隐匿查询方式获取大数据侧标签化信息，确保数据不出域前提下的联合分析。

参考文献：

[1] 周茂雄. 国内外数据安全研究领域前沿动态追踪：基于 2013 年以来的文献计量分析[J].科技管理研究，2022，42(12):17-27.

[2] 侯雨桐,马兆丰,罗守山.基于区块链的数据安全共享与受控分发技术研究与实现[J].信息网络安全，2022，22(02):55-63.

等保 2.0 综合管理平台的设计与实现

陈 明　林志刚　林传捷
（福建医科大学附属第一医院信息中心，福建 福州 350005）

摘　要：本文以某三甲医院为例，提出一种基于等保 2.0 综合管理平台的安全建设方案，实现对信息系统“定级、备案、整改、测评、安全自查和监督检查”的整体过程运行管理，克服医院无法掌握系统安全状态、无法管理等保工作进度等问题，等保管理趋向标准化，有效提升医院安全管理效率和质量，为其它单位等保全流程的建立提供参考。

关键词：等级保护、等保综合管理平台、安全管理

引言

随着信息技术的发展，国家与社会的运转越来越离不开各种各样的信息系统，重要的、基础的信息系统更是关系着国家命脉。信息系统及网络设施的安全性直接关系到医院医疗工作的正常运行，一旦网络瘫痪或数据丢失将会给医院带来巨大的灾难和损失。医院信息系统涉及大量经营和患者医疗等隐私信息，这类信息的泄露和传播将会给医院、社会和患者单来风险[1]。在新的安全形势下，国家越来越重视等级保护工作，现阶段医院对重要等级信息系统开展等级保护工作主要依赖于第三方安全服务机构提供的等级保护相关服务，面对第三方繁杂的报告和文档，其难以真实感受等级保护给重要等级信息系统带来的安全性变化，也无法在第三方安全服务机构提供服务的过程中感受到等级保护工作的进展，更难以从第三方安全服务单位的报告及官方文档中快速获得有关自有重要等级信息系统安全方面的建议[2]。针对上述缺点，本研究以某省属三级甲等医院的等保工作为例，针对医院在等保工作中的困扰，依据国家网络安全等级保护等级保护 2.0 相关标准并进行总体归纳、综合分析后，提出新的一套涵盖安全物理环境、安全通信网络、安全区域边界、安全计算环境、安全管理中心、安全管理制度、安全管理机构、安全人员管理、安全建设管理、安全运维管理等等保相关基本要求的配置检查、管理配套的解决方案，实现对信息系统“定级、备案、整改、测评、安全自查和监督检查”的整体过程运行管理，帮助医院建立等保全流程管理[3]。

1 主要问题

该省属三级甲等医院依托于第三方安全服务机构完成等级保护建设及运维工作，最终完成等级保护测评和等级保护备案，其安全技术防护体系的建设已经趋于完善，但是该医院在等级保护工作过程中仍存在以下问题：

1.1 无法掌握系统的安全状态

通过第三方服务文档，医院无法明确的掌握等级信息系统的总体安全状态，对整体的安全状态没有清晰，完整的认识，无法简单明了地了解等保对象的指标差距、总体情况和具体项目，不能通过这些文档来跟踪相关指标差距项目的实时状态，面对监管部门不定期的安全检查，缺乏全面的安全数据支撑，面对检查时往往是被动状态。

1.2 无法即时管控工作进度

等保建设工作属于体系化、标准化、规范化等要求较高的工作，医院在对等级保护工作内容还未达到透彻了解的情况下，难以做到对等级保护工作进行合理、有力、规范的控制和管理[4]。面对繁杂的报告和文档，医院难以感受等级保护为信息系统

带来的安全性变化，也无法在第三方安全服务机构提供服务的过程中感受到等级保护工作的进展，更难以从第三方安全服务单位的报告及官方文档中快速获得有关自有重要等级信息系统安全方面的建议。

1.3 无法实施等保标准化管理

医院在等级保护建设及运维工作中存在以下问题：各种管理制度不够完善，未进行体系化的梳理与落实；虽通过第三方服务方式进行增补与修订，但医院在等级保护建设及运维的工作中，难以根据本单位的实际情况将制度系统地、规范地应用到信息系统的日常管理之中，并规范有序地实施等级保护的标准化管理。

2 现有解决方案分析

目前国内针对等保管理主要的解决方案为信息安全等级保护综合管理系统和等保检查工具箱。信息安全等级保护综合管理系统适用于政府及企事业单位，为政府及企事业单位日常网络与信息安全管理工作的日常管理支撑工具，可为相关部门网络与信息安全管理工作提供专业化的技术支撑，指导并协助用户开展网络与信息安全日常管理工作，提升工作效率，提高检查质量。这类解决方案适用于大型行业内部的上传下达工作，提高等级保护的日常运维效率，但对等保的资产梳理、人员管理、建设管理等其他方面需求不能充分满足。等保检查工具箱适用于网安部门、测评机构并为其提供专业检查知识和检查方法，提高其标准化、规范化水平，它是公安机关网安部门开展网络安全检查工作的一体化专用便携式监察装备，具有规范检查、工具调用、结果展示等功能，集成定制有专门的安全检查工具，为公安机关网络安全执法检查提供专业检查知识和检查方法，提高网络安全执法检查的常态化、标准化和规范化水平。这类解决方案适用于网安监察部门提供检查效率，对单位本身的等保管理不能提供有力的支撑[5]。

3 方案目标与要求

3.1 方案目标

针对医院在等保工作中遇到的各种问题及现有方案的不足，结合医院等保工作实际情况，本文提出基于等保 2.0 的综合管理方案的解决方案。该方案面向信息系统运营商，旨在帮助医院建立针对各类等保对象建立等保全流程管理，动态掌控等保对象的合规与安全状态，建立体系化、标准化、规范化的管理体系，提高工作效率和安全管理水平。

3.2 技术要求

依据国家等保 2.0 的基本要求以及安全设计技术要求，方案涵盖安全物理环境、安全通信网、安全区域边界、安全计算环境、安全管理中心、安全管理制度、安全管理机构、安全人员管理、安全建设管理、安全运维管理等方面的配置检查、管理配套，实现对信息系统的定级、备案、整改、测评、安全自查和监督检查的整个过程进行整体过程运行管理。

4 等保 2.0 综合管理平台方案设计

4.1 等保全局展示

等保罗盘功能作为创建等保对象的入口，能清楚的了解各等保对象的等保工作运行情况，引导用户进行定级、备案、自评、测评等工作，支持查看每个节点的具体步骤并引导用户进入对应节点进行具体工作，对用户的等保阶段工作起到引导作用。支持对单个的定级对象进行废止、删除、修改以及对等保情况进行复评以及归档等操作。

4.2 系统定级管理

平台定级管理功能可使医院快速了解单位内部各等保对象定级情况，并引导用户进行初步定级和建立定级过程所需的各类材料清单，同时提供定级工作所需的各类材料模板，支持自定义定级相关文档和定级相关材料的自动化输出并进行有效管理，辅助用户在定级过程进行管理。

4.3 系统备案管理

系统备案管理功能帮助用户单位快速了解单位内部各等保对象备案情况，并引导用户进行备案管理和建立备案过程所需的各类材料清单，同时提供备案工作所需的各类材料模板，支持自定义备案相关文档和备案相关材料的自动化填充与输出并进行有效管理，辅助用户在备案过程进行管理。

4.4 等保指标自评

等保指标自评提供等级保护基本要求的指标差距分析与整改向导功能，通过内建在系统中的有关等级保护基本要求合规性的多种指标数据库，如等级保护基本要求指标库、等级保护推荐策略基线库、信息产品安全配置操作知识库等，与等级信息系统资产进行匹配后，构建等级信息系统的等级保护基本要求合规状态属性模型，为等级信息系统的每项资产提供相应的等级保护基本要求合规性推荐指标、操作指引、操作记录及符合性辅助判断等关键功能。

4.5 等级测评

等级测评功能帮助医院快速了解单位内部各等保对象等级测评情况，并引导用户进行等级测评管理和建立测评过程所需的各类材料清单，同时提供测评工作所需的各类材料模板，支持各类自定义测评过程相关文档和测评相关文档的自动填充与输出并进行有效管理辅助用户在等级测评进行管理。

4.6 等保归档

等保归档功能提供各个定级对象各个年度的等保材料归档情况，提供全部下载、按时间搜索按定级对象搜索等功能，并且支持对单个定级对象的定级、备案、自评、测评、管理制度五大维度的单独归档文件管理，方便用户在进行安全检查、等级测评等情况下可以快速调用。

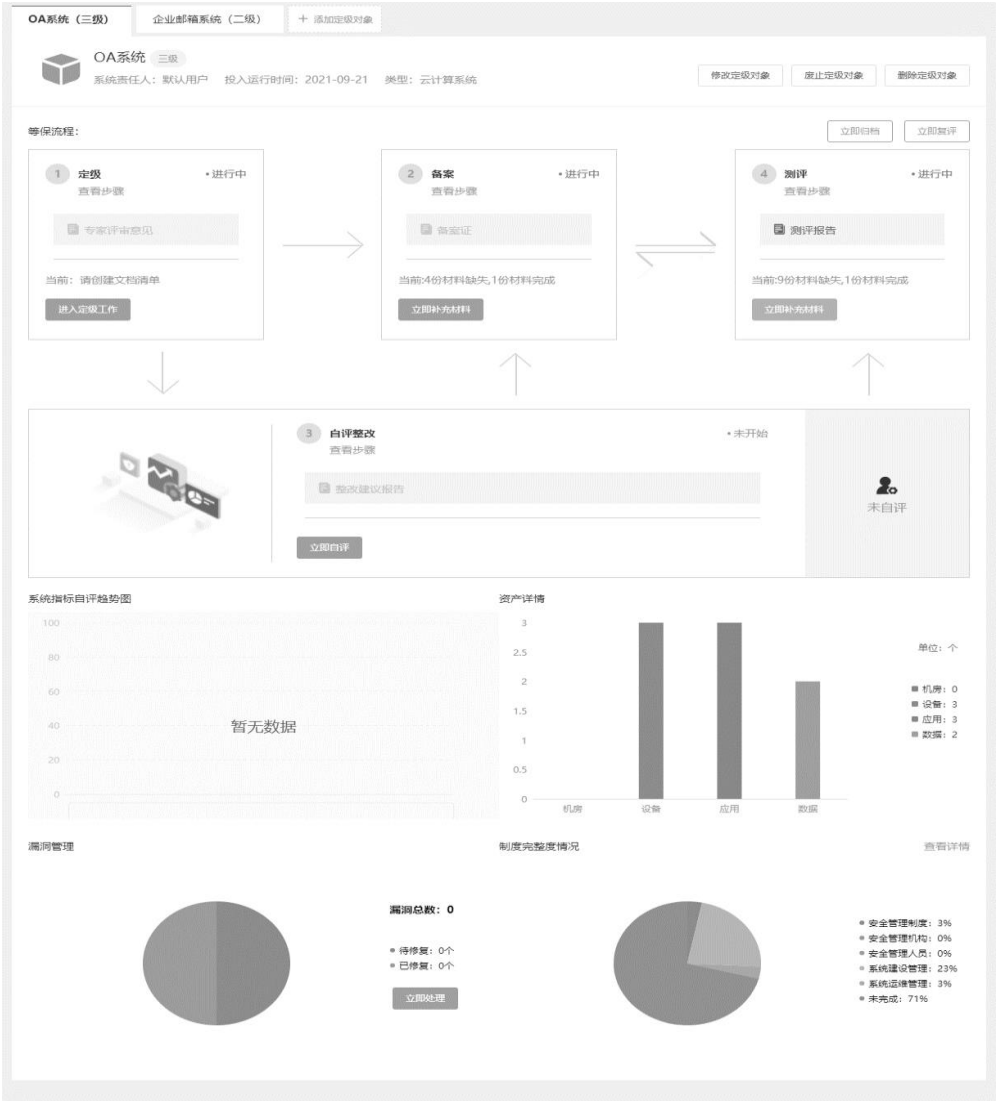


图 5-1 等保 2.0 综合管理平台

5 应用效果

该省属三级甲等医院依据以上方案，开发了等保 2.0 综合管理平台（图 5-1），于 2022 年 3 月开始运行，实现了定级备案、建设整改、等级测评等工作的全周期管理。

5.1 建立并跟踪等级信息系统合规状态

等级保护综合管理平台通过内建等保合规基本要求指标库、等保合规性推荐策略基线库，内容上覆盖了等保合规基本要求中的安全物理环境、安全通信网、安全区域边界、安全计算环境、安全管理中心、安全管理制度、安全管理机构、安全人员管理、安全建设管理等全部指标，通过自动符合度和满足度计算，并结合部分人工辅助判断之后，构建等级信息系统的等保合规基本要求合规状态属性模型，实现医院对信息资产进行实时的跟踪、分析和展示,如图 5-1 所示。

图 5-1 合规自评状态界面

5.2 直观准确地掌控信息系统安全状态

等级保护综合管理平台通过内建的专业漏洞信息库，漏洞库针对多种漏洞库进行内容优化、评级优化和解决建议优化，实现信息系统进行清晰、完整的安全风险综合展示，使医院对信息系统安全状态进行直观、准确、及时、动态地掌握。同时，医院还能获得针对各对象化资产的安全风险进行持续的评估跟踪、风险分析及加固建议、复查对比等涉及等保合规信息系统的安全风险评估加固工作的分析、管理与指导等相关支持，通过与等保合规基本要求相关指标的联合量化匹配，进一步为等级信息系统的对象化资产提供安全性与合规性双融合总体动态追踪展示。

5.3 规范有序实施等保合规标准化管理

等级保护综合管理平台内建符合等级保护安全管理基本要求所包含的安全管理制度、安全管理机构、安全人员管理、安全建设管理等方面的体系及安全管理及制度实践数据库。同时，等级保护综合管理系统配套安全管理制度运行各个环节所涉及的各类数据、信息和材料的综合填报、收集等功能，为用户提供涵盖等保合规所要求的安全管理制度、安全管理机构、安全人员管理、安全建设管理等安全管理要求的制度、规范、流程、记录一体化运行，帮助用户单位面向信息系统实现便捷、高效、标准、规范的在线实时等保合规标准化管理[6]。

6 总结

该省属三级医院在完成等级保护综合管理平台的建设后，其资产管理、等保管理、漏洞管理、制度管理等功能，帮助医院实时跟踪等保信息系统合规状态和等保工作进度，规范有序地实施等级保护标准化管理，提高其等保管理工作效率和整体安全管控能力。

参考文献：

[1] 基于三级等保标准的医院信息安全体系建设实践[J].魏勤.医学信息学杂志，2019，40（2）：35-39.

[2] 唐江波.基于医院信息安全等级保护的整改实践[J].中国数字医学，2018，13（11）：83-86.

[3] 中华人民共和国公安部.信息安全技术网络安全等级保护基本要求:GB／T 22239-2019[S].北京:中国标准出版社,2019.

[4] 陈明.掌上医院平台信息安全风险分析与控制[J].福建医科大学学报(社会科学版),2019,(第 2 期): 22-26,67.

[5] 陈明,林康,林志刚.医院网络公众服务安全问题分析与防护设计[J].中国卫生信息管理杂志,2021,(第 1 期): 106-110.

[6]陈明,林志刚.林传捷，基于态势感知的医院安全实践[J].中国卫生产业,2022,(第 10 期): 102-107.

基于人工智能的恶意加密流量和暗网流量检测解决方案

邹 芳
（中国移动通信集团福建有限公司，福建 福州 350001）

摘 要：当今社会互联网迅速发展的同时，网络安全问题也逐渐成为人们关注的焦点。流量监测作为当前有效分析网络状况的方式之一，人工智能的恶意加密流量和暗网流量一直以来都是网络流量监测的难点与重点，传统的监测方式已无法对其进行有效的监测。通过人工智能检测技术与传统安全检测技术相结合，基于人工智能检测模型对全网实时加密流量进行检测，实现对恶意加密流量和暗网流量的全场景监控，具有极高的可行性和广阔的应用前景。

关键词：暗网、恶意加密、流量检测、人工智能检测

0 背景

近年来为保障通信的安全和隐私，对网络流量加密的企业已超过了 60%。但是无形之中网络流量的加密也在网络安全工作中埋下了新的隐患，据 Gartner（高德纳）公司预测，在 2019 年后，超过 50% 的恶意软件活动，将利用某种类型的加密以隐藏交付、命令、控制活动以及数据泄露。

暗网是深网中的一小部分，是拥有特殊域名的 Web 站点，仅能通过特殊软件、特殊配置进行访问，且使用搜索引擎无法直接对其检索。它的网址与一般的网址不同，以顶级域名后缀“.onion”结尾，且无法通过一般的浏览器对其进行访问，只能通过暗网的浏览器才能访问，暗网用户采用高度加密的方式通信，用户以保密的方式进行文件共享、互相交流、发布博客等操作，同时该方式也极易被用于非法交易、非法论坛以及恐怖分子的介质交流等违法活动中。

随着流量分类技术在信息安全领域的广泛应用，IPv6 网络的扩大以及各种应用的迁移和增多，互联网服务的快速发展和加密技术的广泛应用使其成为一个开放的挑战。近年来，各种增强隐私的工具都采用了加密技术，同时加密流量技术也被黑客所利用，用来进行僵尸网络中对于受控机的 C&C 传输。目前的互联网中的流量根据类型分有 Browsing、Voip、Email、Chat、Streaming、File Transfer、P2P 等七种，例如目前基于 IPv6 环境下的互联网流量中加密流量普遍有 Tor 流量、Shadowsocks 流量和 VPN 流量，准确的检测出这些加密流量对识别网络安全中僵尸网络的具有重大意义。

1 需求分析

1.1 传统安全问题

1. 解密流量是传统处理加密流量问题的方式之一，例如使用新一代防火墙等安全设备来查看流量。采用此类方式不仅耗时长，且还需要在网络之中增设额外的设备，背离了使用加密技术解决数据隐私的初衷。同时不能对无法获取秘钥的加密流量进行解密及检测。

2. 新一代威胁通常使用多种手段并且经过多个阶段来穿透一个网络以窃取信息。攻击者结合使用 Web、电子邮件和基于文件的攻击方式进行攻击。当前的防火墙，IPS，防病毒和 Web 安全网关几乎没有能力阻止使用零日漏洞、一次性恶意软件以及 APT 高级攻击手段的攻击者。

3. 传统的安全技术依赖于静态的基于签名的

或基于列表的模式匹配技术。无法对许多零日和定向型威胁（通过在无辜的网页上或可下载的文件如 JPEG 图片和 PDF 文档里隐藏新型植入恶意软件来渗透系统）进行监测。

4. 通常传统的安全防御措施是将每个攻击方式作为单独的路径，每个阶段作为独立的事件来检查，而不是将这些阶段和方式进行关联，作为一系列精心策划网络事件来检测分析。

1.2 需求分析

1. 能够使用有监督机器学习融合模型和深度学习模型综合检测，通过大量的训练数据训练出可靠的恶意加密流量以及暗网流量检测模型。

2. 能够深度分析和要素统计大量使用加密通信的恶意样本、各类加密通道的攻击行为和多种恶意或非法应用，基于人工智能检测模型对全网实时加密流量进行检测，能够识别恶意加密流量和所属的恶意软件通信类型，能够通过人工智能检测模型对全流量的暗网流量进行检测和分析。

3. 能够结合人工智能检测技术和传统安全检测技术，从多个检测维度构建恶意加密威胁的对抗体系，能够有效识别、检测和防御恶意软件使用加密通讯，加密通道中的恶意攻击行为、恶意或非法加密应用、暗网通讯。

1.3 整体架构

基于 AI 的恶意加密流量和暗网流量检测系统整体系统架构如下图一，监听口接收镜像/分光流量，通过流量采集引擎对数据包进行快速处理以及硬件资源调度。通过筛选引擎过滤不关注的数据，然后二次处理过滤后的数据，将其分别进行特征检测、元数据提取、文件提取、流量存储处理等。通过特征检测引擎来对基于特征的已知攻击进行检测，通过元数据、文件提取实现检测数据预处理，通过流量存储实现数据留存取证。之后通过中间件泛化处理元数据和事件，将处理后的数据提交至 AI 检测引擎、异常行为检测引擎、文件检测引擎、威胁情报检测引擎以及（Yara/JA3/SSL）检测引擎，结合关联引擎进行集中检测，最后将检测结果以日志/告警形式输出展现。

图一 系统架构图

1.4 技术功能模块

基于人工智能的恶意加密流量和暗网流量检测解决方案主要为四个部分，分别为恶意加密流量人工智能检测、Tor 流量人工智能检测、VPN 流量人工智能检测和 ShadowSocks 流量人工智能检测。

恶意加密流量人工智能检测主要是通过提取恶意代码家族的加密网络会话基因特征（DNS 特征、TLS 元数据、HTTP 特征、包特征信息等）训练形成恶意代码加密通讯检测模型。

Tor 流量人工智能检测主要是采用构建 Tor 流量或非 Tor 流量来捕获环境，进行同类应用的数据传输（包括但不局限于浏览器、邮件、聊天工具、视频流、音频流、文件传输、P2P、VoIP 等），从

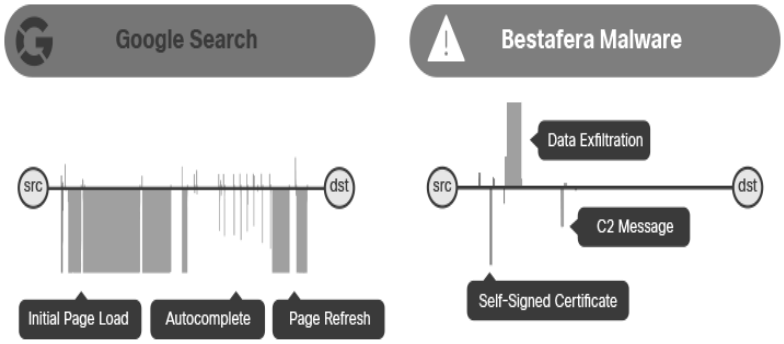
而提取出步态指纹特征数据集训练并建立暗网检测模型。

VPN 流量人工智能检测主要是采用构建流量捕获环境进行同类应用的数据传输（包括但不局限于浏览器、邮件、聊天工具、视频流、音频流、文件传输、P2P、VoIP 等），从而提取出步态指纹特征数据集训练并建立 VPN 流量检测模型。

ShadowSocks 流量人工智能检测主要是构建流量捕获环境进行同类应用的数据传输（包括但不局限于浏览器、邮件、聊天工具、视频流、音频流、文件传输、P2P、VoIP 等），分别对这些应用提取步态指纹特征数据集，通过 Shadowsocks 步态指纹检测结果生成 Shadowsocks 的服务端黑名单和客户端黑名单，经过训练建立 ShadowSocks 流量检测模型。

1.4.1 恶意加密流量人工智能检测

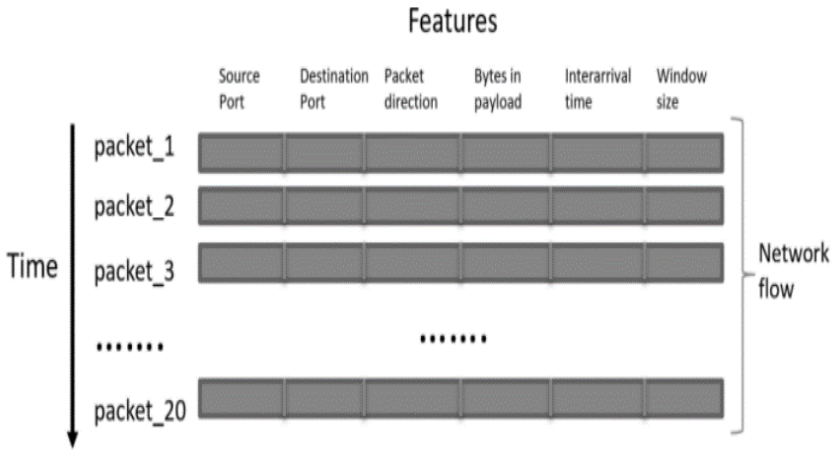
通过提取恶意代码家族的加密网络会话基因特征（DNS 特征、TLS 元数据、HTTP 特征、包特征信息等）训练形成恶意代码加密通讯检测模型，系统获取实时网络会话元数据，构建特征向量，使用检测模型对网络流量进行恶意代码加密通讯检测。如下图二所示：



图二

1.4.2 Tor 流量人工智能检测

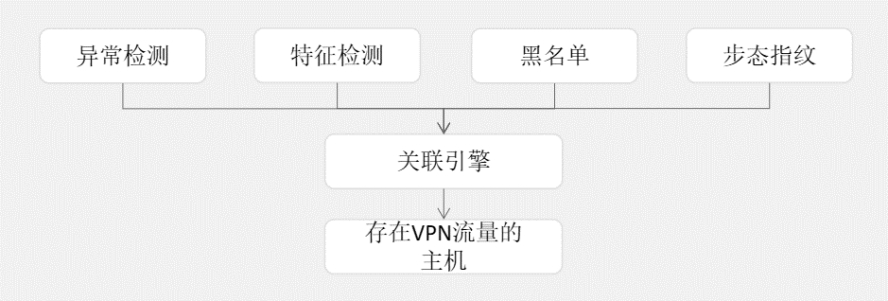
Tor 流量人工智能检测模块主要是采用构建 Tor 流量或非 Tor 流量来捕获环境，进行同类应用的数据传输（包括但不局限于浏览器、邮件、聊天工具、视频流、音频流、文件传输、P2P、VoIP 等），从而提取出步态指纹特征数据集训练并建立暗网检测模型。系统获取实时网络会话元数据，构建实时步态指纹特征，使用暗网检测模型对网络流量进行暗网通讯检测。如下图三所示：



图三

1.4.3 VPN 流量人工智能检测

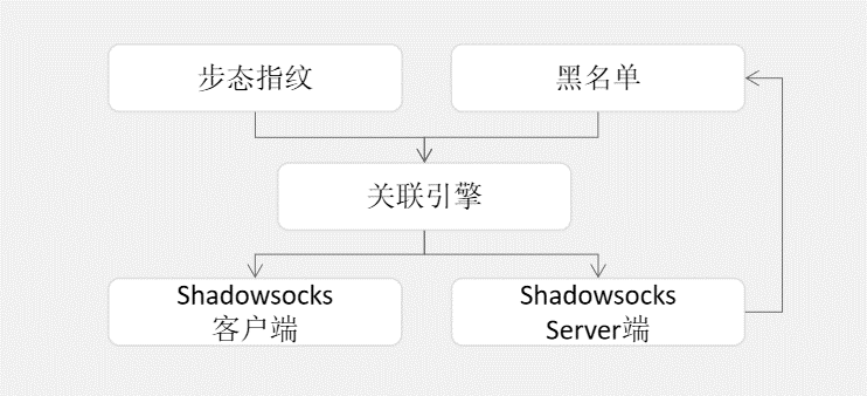
采用构建流量捕获环境进行同类应用之间的数据传输（包括但不局限于浏览器、邮件、聊天工具、视频流、音频流、文件传输、P2P、VoIP 等），分别提取步态指纹特征数据集训练并建立 VPN 流量检测模型，获取实时步态指纹特征，使用 VPN 流量检测模型对网络流量进行 VPN 流量检测。VPN 流量检测模型框架如下图四所示：



图四

1.4.4 ShadowSocks 流量人工智能检测

构建流量捕获环境进行同类应用的数据传输（包括但不局限于浏览器、邮件、聊天工具、视频流、音频流、文件传输、P2P、VoIP 等），分别提取步态指纹特征数据集，通过 Shadowsocks 步态指纹检测结果生成 Shadowsocks 的服务端黑名单和客户端黑名单，经训练建立 ShadowSocks 流量检测模型，获取实时步态指纹特征，使用 ShadowSocks 流量检测模型有效精准的定位存在 ShadowSocks 流量的主机。ShadowSocks 流量检测模型框架如下图五所示：



图五

获取实时步态指纹特征，使用 ShadowSocks 流量检测模型有效精准的定位存在 ShadowSocks 流量的主机。

2 方案创新能力

2.1 基于数据流上包统计特征构建模型能力

我们一般将模型评价分为两类：功能指标、性能指标。功能指标用于评价模型的识别效果，性能指标用于对评价模型的识别效率。在本次实验中，使用精确率（Precession）、召回率（Recall）、准确率（Accuracy）以及 F1 值（F1−socre）等功能指标，用以衡量模型的检测能力。

准确率（Accuracy）是指分类正确的个数占总样本的比例，用于衡量分类器作出的判决中总体的正确率情况，公式如下：

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

召回率（Recall）是指正样本被分类正确的个数占总正样本数的比例，用来衡量模型对正样本的检测能力，公式如下：

$$R = \frac{TP}{TP + FN}$$

精确率（Precession）是指分类为正样本实际为正样本的比例，用来衡量对正样本结果中的预测准确度，公式如下：

$$P = \frac{TP}{TP + FP}$$

F1 值（F1−socre）兼顾分类模型的精确率和召回率，可以看作是模型精确率和召回率的一种加权平均，对于不平衡样本集 F1 值更能衡量模型的分类能力，公式如下：

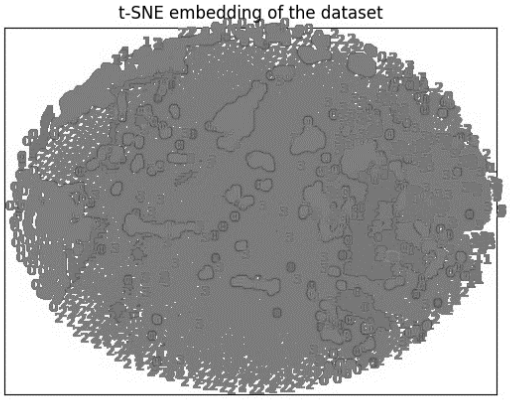
$$F1 = \frac{2 * P * R}{P + R}$$

基于数据流上包的统计特征，提取特征向量后所得到的训练样本，如表一所示：

表一 数据流上包统计特征构建的训练样本

样本类型	样本总数（单位：个）
白样本	31161
Tor 直连样本	45472
Shadowsocks 代理样本	225596
VPN 代理样本	9723

以上样本的数据降维分布，如图六所示：



图六 数据集 t−SNE 分布图

在数据集的模型训练中，选择目前较流行的分类算法：XGBoost、LightGBM、RandForest 和 Logistic Regression 进行模型训练，且全部采用算法模型的默认参数进行训练。在数据集的划分中，训练集:验证集:测试集按照 4:2:2 的比例来进行划分，并在训练中采用 10 折交叉验证来进行模型训练和结果评估。

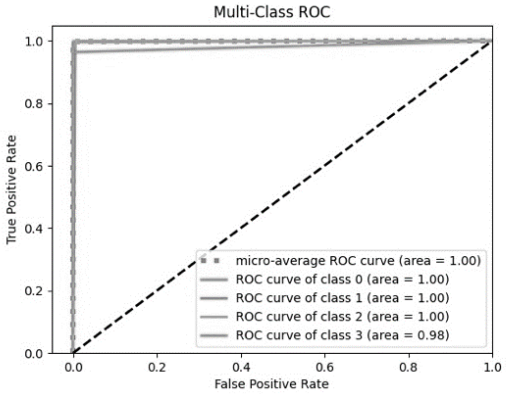
基于数据流上包的统计特征对各模型进行训练后，在测试集上得出的评价指标如表二所示：

表二 不同算法模型在测试集上的评价结果 1

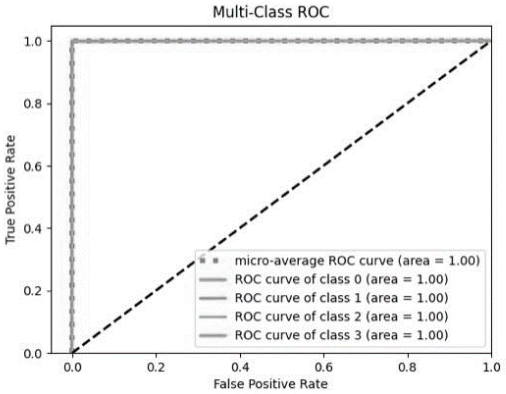
分类模型	Accuracy	Recall	Precision	F1−score
XGBoost	0.99983076	0.99582853	0.99963828	0.99771797
LightGBM	0.99981829	0.99350040	0.99835846	0.99590395
Random Forest	0.99979157	0.99414401	0.99969957	0.99688789
Logistic Regression	0.98330245	0.87101271	0.95909096	0.90579628

从结果中可以看到，XGBoost、LightGBM 以及 Random Forest 中的各功能的评价指标都达到理想效果，而 Logistic Regression 则相对较差，其中 Recall 仅有 0.87，表明该模型对 Tor 流量的检测能力不如另外三个模型。

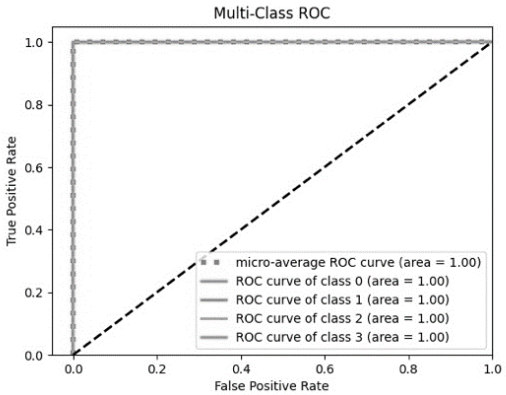
为进一步准确评价模型，将各个模型的 ROC 曲线图以及混淆矩图进行展示，如图七和图八所示，class 0: 白样本，class 1:Tor 直连样本，class 2:Shadowsocks 代理样本，class 3: VPN 代理样本。



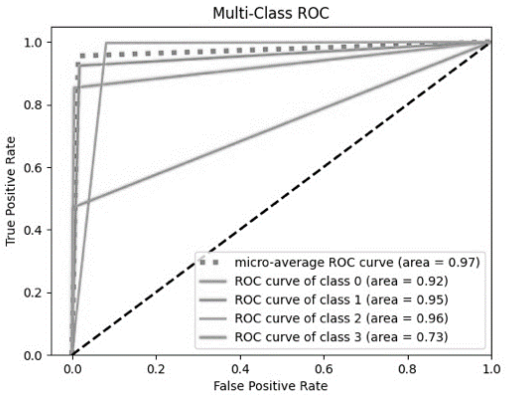
XGBoost



LightGBM

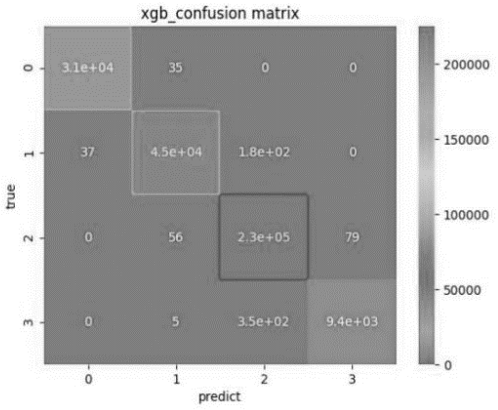


Random Forest

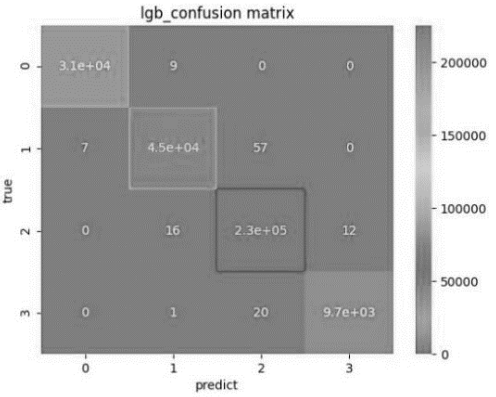


Logistic Regression

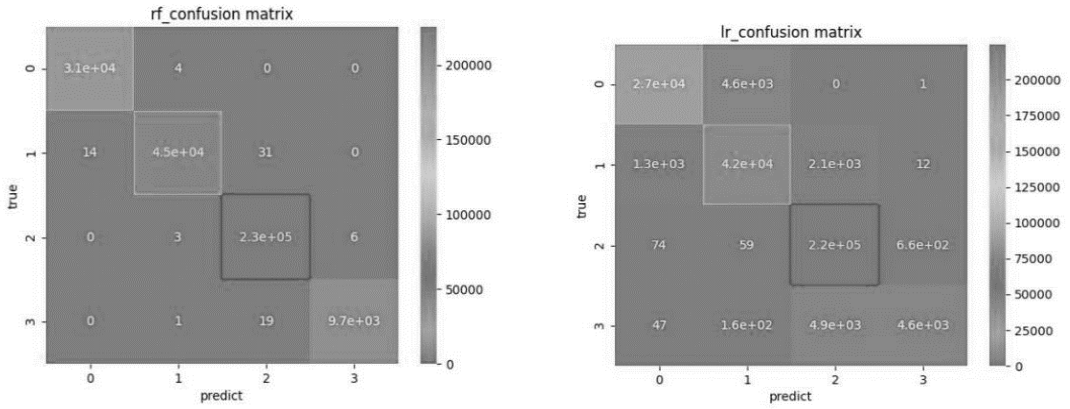
图七 模型 ROC 曲线图



XGBoost



LightGBM



Random Forest　　Logistic Regression

图八 模型 ROC 混淆矩阵图

从表二：指标评价结果表综合 ROC 曲线以及混淆矩阵图来看，XGBoost、LightGBm 以及 Random Forest 在 Tor 流量的综合检出能力及各类别的检出能力上都有较好的表现，而 Logistic Regression 产生误报的概率更大。

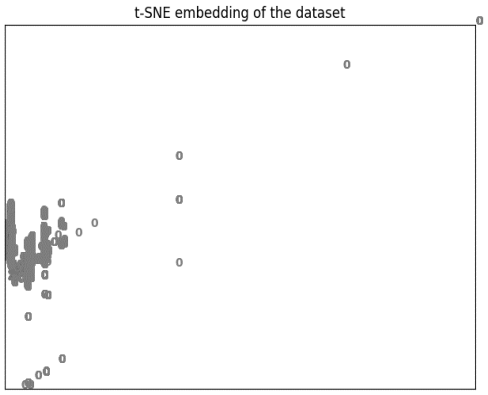
2.2 基于加密协议 TLS/SSL 协议构建模型能力

基于加密应用协议 TLS/SSL 来进行特征提取，不仅利用了加密协议 TLS/SSL 通信特征，而且利用了数据流上包的统计特征，并对 DNS 上下文统计特征进行关联，从而构建 AI 训练的特征工程。

表四 DNS+TCP/UDP+SSL 训练样本

样本类别	样本数量
白样本	194952
Tor	12035
shadowsockets	8021

以上数据的降维分布，如图九所示：



图九 样本 t-SNE 数据分布图

在该数据集的模型训练中，选择目前较为流行的分类算法：XGBoost、LightGBM、RandForest 以及 Logistic Regression 进行模型训练，且训练全部采用算法模型的默认参数进行。在数据集的划分中，训练集:验证集:测试集按照 4：2：2 的比例进行划分，并采用 10 折交叉验证在训练过程中进行模型训练和结果评估。

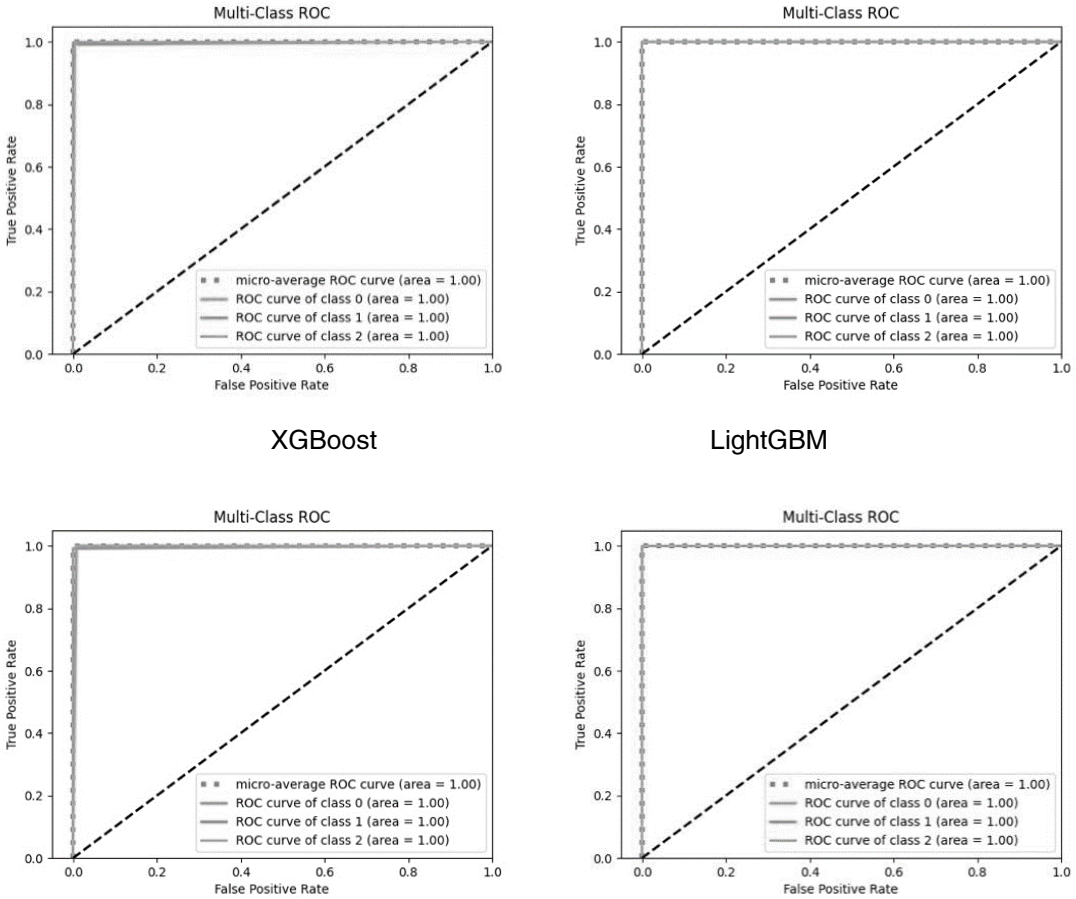
训练后的各模型在测试集上得出的评价指标，如表五所示：

表五 不同算法模型在测试集上的评价结果 2

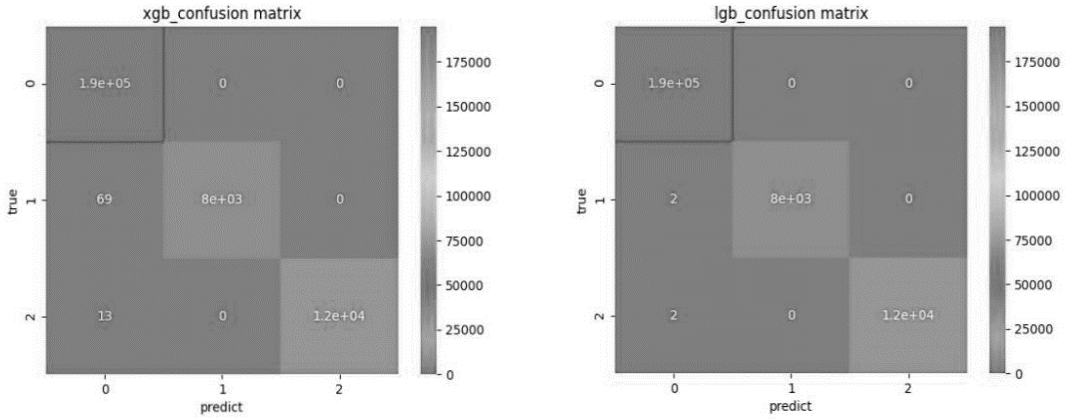
分类模型	Accuracy	Recall	Precision	F1-score
XGBoost	0.99997861	0.99977728	0.99999229	0.99988475

LightGBM	0.99997861	0.99977728	0.99999230	0.99988474
Random Forest	0.99993346	0.99935500	0.99997604	0.99966532
Logistic Regression	0.99973386	0.99792784	0.99954118	0.99873235

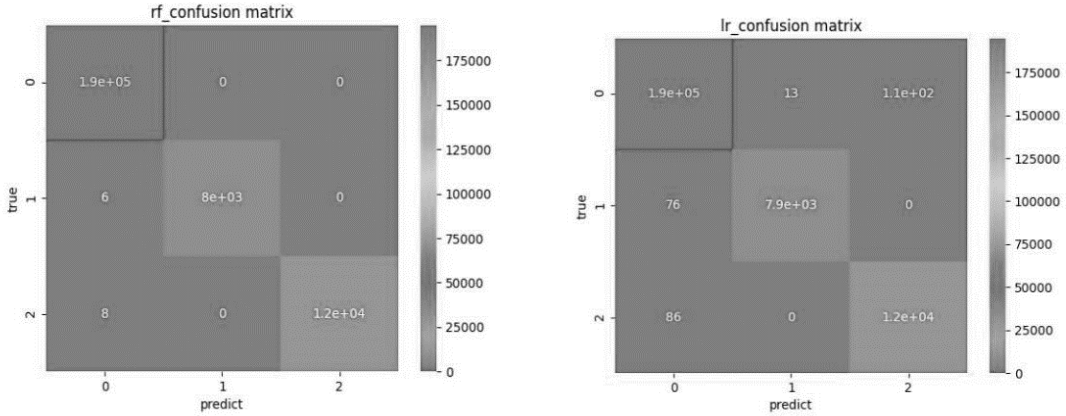
从结果中可以看到，各个模型在各个功能评价指标上表现都较好。为了对模型进行更加准确的评价，将各个模型的 ROC 曲线图以及混淆矩图进行综合展示，如图十和图十一中所示。class 0: 白样本，class 1：Tor 直连，class 2：Shadowsocks。



图十　四种模型的 ROC 曲线



XGBoost. LightGBM



Random Forest

图十一 四种模型混淆矩阵图

从表五模型评价指标结果以及图十的 ROC 曲线以及图十一的混淆矩阵图中可以看到，各模型在数据集上的表现都较为稳定，且检测能力较好。

在实验环境抓取的模型训练数据，除了对这些实验数据进行交叉验证之外，同样采用 CIC Tor2016 公开数据集验证模型的检测效果。

3 方案价值

为实现对网络中良性/恶意的流量以及恶意威胁检测的目标的实时区分，深入研究恶意加密流程的特征，并通过机器学习的方式，对正常流量和恶意加密流量的特征进行学习，从而达到实时区分和检测恶意威胁的目的。对日常流量抽丝剥茧，将种类繁杂的普通加密流量与特殊的“暗网”相关流量进行区分，发现其中的“暗网”访问者、“暗网”服务中继节点及“暗网”相关服务提供者，实现对“暗网”流量的全场景监管，发掘有效“暗网”事件线索，具有极大的推广和应用价值。

一种基于机器学习算法的僵尸网络检测方法

余建[1] 李宗铖[1] 林志兴[1] 郎旭[2]
（1.三明学院 网络中心，福建 三明 365004；
2.福建中信网安信息科技有限公司，福建 福州 350011）

摘 要：当前检测 Fast-flux 僵尸网络流量的方法主要分为主动与被动两类，前者对设备的网络带宽要求较高，后者可能存在特征提取流程复杂的难点。因此，针对僵尸网络的特点，给出一种机器学习的 Fast-flux 僵尸网络流量的方法。首先将 DNS 报文数据流转换为灰度图像，再使用时序卷积网络僵尸网络流量检测模型进行训练，经过分类器得到结果。最后使用 ISOT 2010 公开数据集与自收集数据集进行实验后，实验结果表明，采用该方法的实验结果准确率达到了 94.9%，召回率达到了 95.3%，精确率达到了 94.4%。

关键词：时序卷积网络；僵尸网络；Fast-flux

0 引言

近年来，互联网长期行驶在发展的“快车道”，为人民的生活带来了颇多便利。但是随着网络技术的迅猛发展，不法分子也紧跟着时代的潮流，利用网络技术来获取非法利益，带来了诸多安全问题，例如，如对计算机资源（CPU、内存、存储、网络等资源）的窃取、盗用、损坏等恶意操作。不法分子利用了包括但不限于病毒、木马、蠕虫等网络技术，并且不法分子多以僵尸网络为基础，操控其对个人或者组织进行信息窃取、挖矿、勒索等恶意行为。由此可见，网络安全领域也正受到愈发严峻的挑战，而僵尸网络即是该领域所面临的严峻挑战之一。

僵尸网络（Botnet）是指攻击者采用一种或多种传播手段，将恶意程序如计算机病毒、网络蠕虫、特洛伊木马和后门工具等植入主机，并进行感染，从而控制大量主机，然后通过一对多的命令与控制信道对主机进行控制，从而形成僵尸网络[1]。近些年，我国的经济高速发展，人民的生活水平日益提高，使得智能手机、计算机、智能家居等设备的人均持有量大有提升，而现在这些设备也能够被僵尸网络控制并加入到僵尸网络中。设备感染僵尸程序后，攻击者通常会使用 C&C 服务器来控制，从而达到攻击者的预期目标。

根据国家应急互联网响应中心（CNCERT）发布的《2020 年中国互联网网络安全报告》[2]显示，2020 年，我国境内 C&C 服务器的 IP 地址有 12,810 个，我国境内共有近 534 万个主机被植入僵尸程序。依据知名网络安全公司 Symantec 发布的《2019 年互联网安全威胁报告》[3]显示，2018 年，曾经感染了全球约 160 万台设备，被称为最危险的恶的意软件和过去十年最具破坏性的僵尸网络之一的 Emotet，继续积极扩大其势力范围。而且 Emotet 还同时被用于传播危险性十足 Qakbot 程序。网络安全行业龙头公司 Sophos 曾指出 Qakbot 程序将使得僵尸网络变得更加复杂，检测难度提升。根据全球网络安全解决方案提供商 Fortinet 发布的 2021 年下半年《全球威胁态势报告》[4]表示，僵尸网络不再以 DDoS 的单一体攻击为主，攻击者转而利用捆绑勒索软件在内的多目标攻击工具等更为复杂的攻击技术。例如，Mirai 程序利用物联网设备构成的 IOT 僵尸网络，导致了大规模互联网中断事故，造成了极其恶劣的影响。

随着互联网的应用与技术的不断进步，僵尸网络也有了新的发展，一些新的技术与机制被引入其中，使得僵尸网络越发的健壮，这给网络安全的防

御人员带来了新的挑战[5]，其中最引人关注的技术就是 Fast-flux 技术。Fast-flux 技术[6]目的是不断变换 IP 地址到同一个域名的映射关系。这种技术是基于 DNS 协议实现的，使用了该技术的僵尸网络的鲁棒性得到了极大地提升。与此同时，检测与防护这类僵尸网络的难度也大幅提升。

由此可见，Fast-flux 等隐蔽技术在僵尸网络中的巧妙应用，使得其在当今的网络态势下，仍然保持着极高的隐蔽性和较高的存活率，这对互联网来说是一个巨大的安全隐患。因此，深入了解僵尸网络的运行机制与发展态势，进一步分析研究 Fast-flux 技术原理，并提出一种不依赖于特征值提取、网络设备负担不大、检测僵尸网络流量准确的检测方法具有深刻意义。

1 僵尸网络原理

僵尸网络（Botnet）是指攻击者通过一些手段传播并利用僵尸病毒，致使许多主机感染而成为受控主机，从而构成一个一对多的可控网络。Bot 原本是指机器人程序，它能够根据事先定义好的规则来批量地处理文件。而在僵尸网络中，bot 指的是那些如同僵尸一样的受控主机。

如今的僵尸网络已经是重大的网络安全隐患之一。由于僵尸网络中的受控主机能够根据控制主机发出的命令进行协同合作对指定设备发起恶意行为，这正是僵尸网络与其他恶意攻击方式最主要的差别。

1.1 僵尸网络的组织结构分析

僵尸网络的组织结构主要分为集中式、P2P 式以及混合式等结构。

（1） 集中式

集中式结构类似于 C/S 结构，受控主机与 C&C 服务器进行通信，接收并执行来自 C&C 服务器的命令。集中式结构的僵尸网络具有反应速度快与协作性良好的优点，但是对 C&C 服务器的可用性依赖程度高。因此，目前的僵尸网络会使用加密混淆技术将通信内容加密，从而提高集中式僵尸网络的生命周期。新型的僵尸网络大都使用了隐匿技术来提高僵尸网络的健壮性，Fast-flux 技术便是常用的隐匿技术之一，也是本文研究的重点。

（2）P2P 式

由于集中式的僵尸网络常常出现因中心控制节点宕机而导致整个僵尸网络无法使用的问题，遂提出了 P2P 式的僵尸网络。在这种结构中，受控主机既可以充当客户端使用，又可以充当服务器使用，这样直接规避了中心控制节点宕机从而产生的影响。

（3）混合式

混合式结构就是以上两种类似的融合使用，但是融合也有主次之分。以集中式为主的混合式结构，从宏观角度看属于集中式。但从微观角度看 C&C 服务器节点部分又是由 P2P 式组成。以 P2P 式为主的混合式结构则存在多个局部中心的 C&C 服务器节点，并且各个 C&C 服务器节点相互联通，最终组成整个僵尸网络。

1.2. 僵尸网络的主要特征

依照僵尸网络的定义与组织结构，可以总结出以下几点特征。

（1）僵尸网络是由控制主机与受控主机构成的远程可控网络。仅由受控主机构成的网络，并不能够称之为僵尸网络。只有当攻击者能够通过僵尸病毒来控制大量主机的时候，僵尸网络才会具有巨大的破坏力。

（2）僵尸网络中的受控主机之间存在相互协作的能力。这种能力就是僵尸网络与其他恶意程序的主要差异之处。计算机病毒、蠕虫、木马、间谍软件等恶意程序主要以破坏本机系统与窃取隐私信息为主，基本不存在相互协作的能力，而相互协作能力却是极大的提升了僵尸网络的破坏性。

（3）僵尸网络的主要目的是为了实施恶意活动。在一些正常服务的网络架构之中也有运用类似僵尸网络这样的组织架构。然而僵尸网络能够成为目前重大的网络安全隐患之一就是由于构建僵尸网络是为了实施恶意活动。僵尸网络的恶意活动包括但不限于窃取秘密、滥用资源、DDoS 攻击以及近期较为严重的虚拟货币挖矿等。

1.3 Fast-flux 技术

为了绕过安全设备的检测与追踪，攻击者会使用各类隐匿技术来加强僵尸网络，比如 Fast-flux 技术。Fast-flux 技术[6]是指不断改变域名和 IP 地址映射关系的一种技术。使用这个技术，能够使大量的受控主机组成一个动态的代理网络，将其背后的 C&C 服务器很好地隐匿起来。

Fast-flux 技术的应用方式主要有两种：Single-flux 模式和 Double-flux 模式。

（1）Single-flux 模式

Single-Flux 模式指的是仅使用了一层 Fast-flux 技术的模式。在 Single-Flux 模式中，一个域名拥有一个 IP 地址池，这个地址池中可能包含几百上千个 IP 地址。攻击者利用自己所拥有的 DNS 服务器以及 IP 地址池作为基础，通过这个 DNS 服务器来实现短时间内多次变换的 C&C 服务器的 IP 地址的目的。

Single-flux 模式下 DNS 解析流程如图 1 所示。为了得到 www.test.com 对应的 IP 地址，僵尸计算机向 DNS 服务器发起查询请求。DNS 服务器将 IP 地址池中的 1.2.3.4 响应给僵尸计算机，同时在响应报文中将 TTL 设置为一个很小的数值。僵尸计算机访问 IP 地址为 1.2.3.4 的服务器进行获取信息。IP 地址为 1.2.3.4 的服务器将相应的信息应答给僵尸计算机。短时间内僵尸计算机想要二次请求 www.test.com 的信息时，由于之前应答设置的 TTL 过小，僵尸计算机本地缓存的记录已经过期。僵尸计算机必须再度对 DNS 服务器请求 www.test.com 所对应的 IP 地址。此次，DNS 服务器从 IP 地址池中选择了另一个 IP 地址 5.6.7.8 返回给僵尸计算机，仍然设置一个很小的 TTL。僵尸计算机访问 IP 地址为 5.6.7.8 的服务器进行获取内容。IP 地址为 5.6.7.8 的服务器响应僵尸计算机的请求，返回相应的内容。

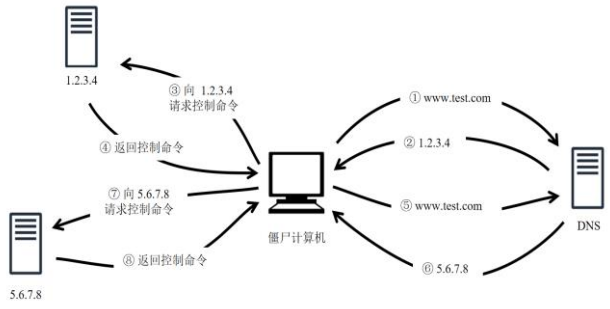


图 1 Single-flux 模式下 DNS 解析流程

（2）Double-flux 模式

Double-Flux 模式指的是使用了嵌套使用了 Fast-flux 技术的模式，简单来说就是在 Single-flux 模式外面套了一层 Fast-flux 技术。在 Double-Flux 模式中，攻击者会构建多个自己拥有的内层 DNS 服务器用于解析 C&C 服务器的域名。并且攻击者会不断改变外层 DNS 服务器上映射到内层 DNS 服务器的 IP 地址的资源记录。这样映射 C&C 服务器域名的内层 DNS 服务器也能够进行灵活变换。由于外层 DNS 服务器并不是攻击者所拥有的。因此，为了规避僵尸网络暴露，顶层 DNS 服务器的域名与 IP 地址映射关系修改频率会比底层 DNS 服务器低很多。Double-flux 模式的运行流程如图 2 所示。

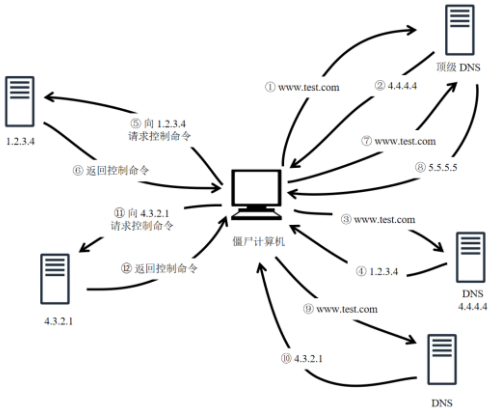


图 2 Double-flux 模式运行流程

2 基于时序卷积网络的僵尸网络流量检测方法研究

通过剖析使用了 Fast-flux 隐匿技术的僵尸网络的底层原理，本文给出一种基于时序卷积网络的僵尸网络流量检测方法。如今虽然众多学者都提出了基于 DNS 流量进行检测恶意的僵尸网络流量的方法，但是基本都偏向使用多个模型融合或者采用人工提取特征等操作。由于多个模型融合的方法通常都会占用大量的系统资源等弊端，以及运用了人工提取特征的方法都可能有提取步骤复杂以及特征有效性参差不齐等弊端，所以本文提出一种利用单一模型且同时规避了人工提取特征这一痛点的检测方法。该方法主要利用时序卷积网络模型对 Fast-flux 僵尸网络流量的特征进行深度发掘，期望提升 Fast-flux 僵尸网络流量检测的有效性。

2.1 检测方法总体架构

本文提出的检测方法总体架构如图 3 所示。

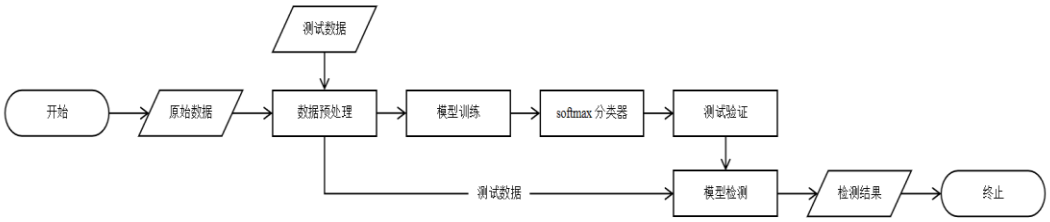


图 3 检测方法总体架构

检测方法的总体架构主要包括以下 4 个环节。

（1）采集流量数据：这个环节主要是采集流量数据，然后对这些数据进行初步的筛选得到所需的部分。为了降低采集工具在采集过程中对设备系统资源的长时占用，本文使用流量镜像将流量引导至指定位置进行采集。之后再将这些数据根据本文检测方法所需的 DNS 协议进行过滤，得到所需的 DNS 流量数据。

（2）数据预处理：这个环节主要是将过滤得到的 DNS 流量提取出 DNS 报文数据部分，再将其转换为模型训练所需的输入格式。本文主要通过编写 Python 脚本对流量进行裁剪、统一化，将原本的流量数据截取出 DNS 报文部分，再统一化尺寸，最后转换成模型所需的图片格式。

（3）模型训练：这个环节主要是调整模型的超参数并训练模型。将上个环节得到的数据集划分出训练集后，使用时序卷积网络模型对训练集进行训练。本文利用单一变量法对模型进行多次调整超参数，最后保留检测性能较好的超参数。

（4）模型检测：这个环节主要是利用上个环节得到的模型进行检测。根据模型中分类器输出的检测结果，可以将进行检测的流量分为正常网络流量和恶意网络流量。恶意流量包含了僵尸网络流量的部分，在此并不对僵尸网络流量再次分类，而是归于同一类。从而锁定僵尸网络的域名，之后就可以对其进行溯源、封禁等其他操作。

2.2 时序卷积网络僵尸网络流量检测模型

时序卷积网络僵尸网络流量检测模型主要由多个时序模块与分类模块顺序串联而成，如图 4 所示。将多个时序模块顺序串联在一起，能够使模型学习到数据之间更深层次的联系。分类模块用于前置产出结果的分类。

（1）时序模块是由膨胀因果卷积、激活函数 ReLU、Dropout 函数以及残差连接组成。首先，数据进入模型的膨胀因果卷积进行学习，接着通过激活函数 ReLU 提升模型的表达能力，然后通过 Dropout 函数临时舍弃部分神经元降低过拟合的可能性。再重复前面的三个操作一次，最后将前面得到的结果进行残差连接后，经过激活函数 ReLU。时序模块的操作流程具体如图 5 所示。其中对于每个膨胀因果卷积，还加入了 Weight Norm 来正则化网络。

（2）分类模块是将时序模块产出的结果经过全连接层，再通过 log_softmax 分类器判断该流量是正常网络流量还是僵尸网络流量。

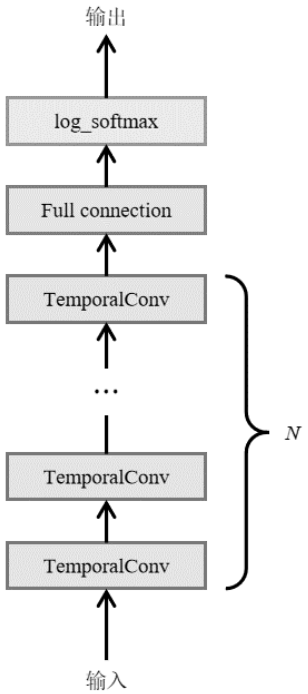


图 4 时序卷积网络僵尸网络流量检测模型架构

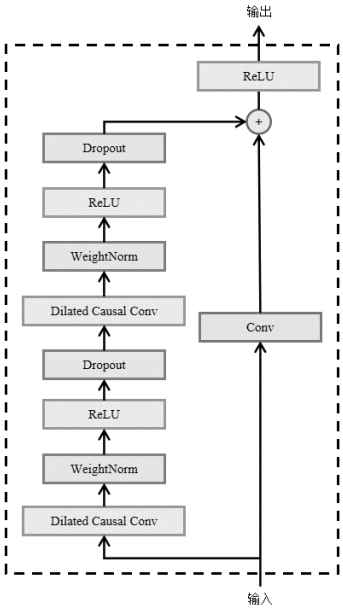


图 5 时序模块流程

根据 Fast-flux 僵尸网络流量图像化后存在相似特征、上下文存在关联性以及周期性的特点，本文提出使用时序卷积网络模型进行 Fast-flux 僵尸网络流量的检测。首先，该模型是以卷积网络为基础，而众所周知，卷积网络能够很好的提取图像的特征。其次，该模型使用因果卷积能够更好的提取流量上下文关联与周期性的特征。目前，大部分学者都使用传统的机器学习或者多种模型混合的方法来最大程度提升分类的准确性，与本文的期望有所不同。所以本文仅用单一模型进行训练，期望在降低模型复杂度与训练难度的同时，使分类的精度依旧保持在较高水平。

2.3 时序卷积网络模型训练

时序卷积网络（Temporal Convolutional Network，TCN）是一种能够直接利用卷积网络的强大特性，且能跨时间步提取特征的模型。时序卷积网络模型的训练速度相较于循环神经网络模型更快。时序卷积网络模型主要由因果卷积、膨胀卷积和残差连接三部分组成。

（1）因果卷积

图 6 为因果卷积示意图。从图中可以看出，在因果卷积中，对于当前层次某个时刻的值来说，它只能够被低于当前层次的那个时刻与其之前时刻的值影响。

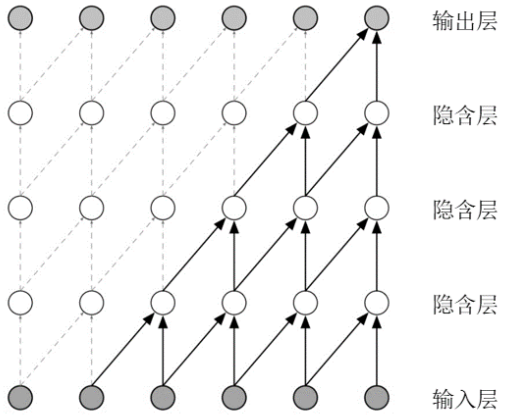


图 6 因果卷积

因果卷积是一个单向的结构，只有存在前面的因，才会产生后面的果。结合 Fast-flux 僵尸网络的特点，因果卷积相较于其他比较传统的方法有更好的上下文关联性。这个过程可以用如下方程式表示。

$$x_l = H_l([x_0, x_1, \cdots, x_{l-1}]) \quad (1)$$

但是，如果因果卷积要能够应对长历史信息的问题，那么卷积层数就必须变多。但是，卷积层数变多又会产生梯度消失、训练复杂度高以及拟合效果差等弊端。为此，时序卷积网络模型使用膨胀卷积来消除这类影响。

（2）膨胀卷积

膨胀卷积能够使模型在维持总体参数个数不变的条件下，提升卷积核的感受野，使得每个卷积输出都覆盖了较大范围的输入。与此同时，又能保持输出特征的尺寸与维度不变。如图 7 所示。

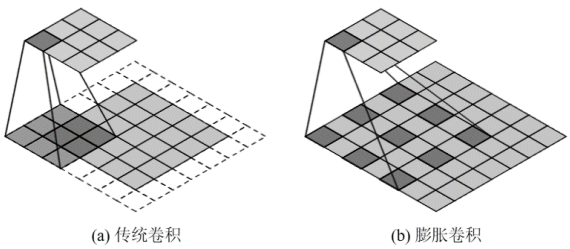


图 7 传统卷积与膨胀卷积的对比

在本文中，使用膨胀系数为 2 的指数次方

（*d=1,2,4,8,16,...*）。将因果卷积与膨胀卷积融合为膨胀因果卷积，如图 8 所示。膨胀因果卷积不仅能够应对长历史的问题，还具备因果性。结合 Fast-flux 僵尸网络的特点，膨胀因果卷积相较于因果卷积能够更好的检测周期性这一特征。

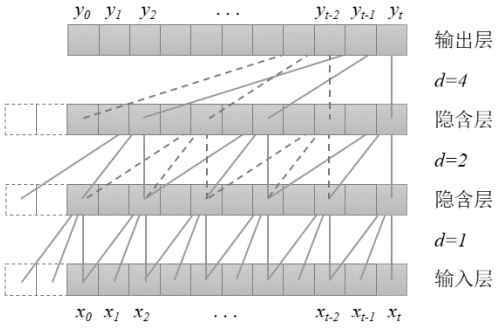


图 8 膨胀因果卷积

（3）残差连接

残差连接在某些方面能够抑制深层网络训练存在梯度消失或爆炸的问题，原因是它能够将原始信息不受影响跨层级传递，如图 3-9 中虚线所示。因此，残差连接在大量实验中被证实是训练深层次网络的重要途径。

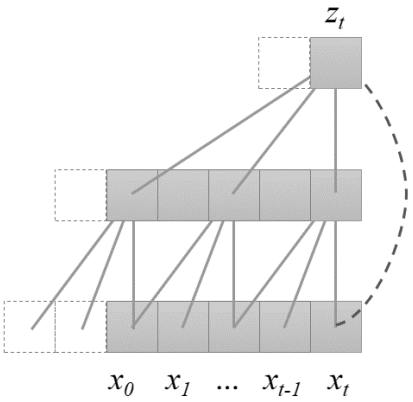


图 9 残差连接

本文之所以选择使用时序卷积网络模型，不仅是该模型比循环神经网络模型训练更加高效，还因为该模型能够较为良好的利用 Fast-flux 僵尸网络的特征。由于 Fast-flux 僵尸网络访问的 IP 地址是不断变化的，但是实现 IP 地址能够发生变化是需要控制主机通知 Fast-flux 僵尸网络的傀儡机以及各种组件，即控制主机发出某条指令后，访问的 IP 地址才会随之变化。如果该 Fast-flux 僵尸网络使用的是预先定义好的变换规则，以某种循环使用 IP 地址池中 IP 地址，即按照一定的顺序访问这些 IP 地址的主机，那么捕获到的 DNS 流量数据也将存在上下文关联性，不会像一般人访问网站产生的流量那样无序。

2.4 分类检测

（1）分类器

本文使用的是 log_softmax 分类器，该分类器是目前较为主流的分类器之一。相比与 softmax 分类器，log_softmax 分类器理论上就是对 softmax 分类器的结果进行对数运算，但实际上 log_softmax 分类器不仅能够加快运算的速度，还能保持数值的稳定性。对于一个数组 *S* 来说，其中的每个元素的 log_softmax 值为

$$\log\_softmax(C_i) = \log_e[\frac{e^{zi}}{\sum_{j=1}^{n} e^j}], i = 1,2,\cdots,N \quad (3-2)$$

其中，C_i 表示模型对输入 *x* 在第 *i* 个类型的评分结果。本文最后的输出结果为两个类别。

（2）损失函数

本文使用的损失函数为负对数似然损失函数（Negative Log Likelihood, NLL Loss），用来判断模型的鲁棒性优劣的问题。NLL_Loss 公式如下。

$$nll\_loss = -\frac{1}{N}\sum_{k=1}^{N} y_k\,(log\_softmax) \quad (3-3)$$

其中，*N* 为样本个数；*yk* 表示第 *k* 个输入数据 *x* 的真实标签；log_softmax 表示输入数据 *x* 的 log softmax 值。

2.5 数据预处理

本文采集到的 DNS 流量都是以 pcap 格式的文件进行存储，而模型所需的输入格式为图像格式。因此需要对 pcap 文件进行处理，转换为模型需要的特定格式。本文使用的时序卷积网络模型是基于卷积网络模型的，所以将 pcap 文件中的流量数据转换为图像格式即可，具体预处理流程如图 10。

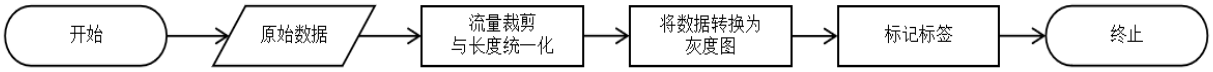


图 10 预处理流程

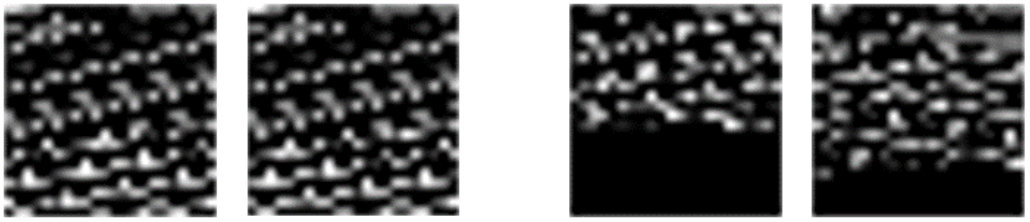
本文提出的检测方法主要针对 DNS 响应包中的 DNS 报文内容进行检测。DNS 报文内容包含了 DNS 响应包中绝大部分有效信息，也是图像化后特征最为明显的部分。仅使用 DNS 响应报文进行训练，虽然损失了部分信息，但是能够减少输入的数据量，也能够提升训练的速度与效率。所以，本文将 DNS 响应包中 DNS 协议数据部分提取出来进行后续处理。

由于本文使用的模型输入格式为图像，需要将数据转换为相同尺寸的图像。所以，本文将对提取出的 DNS 协议数据部分进行统一化处理。本文截取每条数据的前 484B（22*22）的数据，假如某个 DNS 响应包中 DNS 协议数据部分长度不够 484B 时，会使用 0x00 在其末尾进行补足。

在 DNS 报文数据裁剪与长度统一化后，将其转换为 8 位灰度图像即可。对于裁剪并统一化后的 DNS 报文数据按照字节逐个转换成 8 位的灰度像素。将 DNS 报文数据转换为 8 位灰度图后，这个灰度图可以看作 DNS 报文数据的“黑白照”，示例如图 3-5 所示。

从图 11 能够看出，僵尸网络流量数据转换成“黑白照”后重合率较高。而正常的网络流量数据转换成“黑白照”之间差异较为明显。因此，本文推断卷积神经网络分类图像的方法对流量转化成的灰度图进行检测与分类是有效且可行的。

数据预处理完成后，就可以对模型进行训练。



（a）僵尸网络流量　　（b）正常网络流量

图 11 僵尸网络流量与正常网络流量的可视化结果

从以上中能够得出 Fast-flux 僵尸网络流量在 8 位灰度图上呈现出了显而易见的特征。这种将流量数据转化成图片，训练得到的特征暂且定义为空间上的特征，它能够体现出流量数据自身的一些特征。而仅利用了 Fast-flux 僵尸网络流量自身的特征是不够充分的，还需要学习时间维度上的特征，包括周期性、上下文关联性等。因此本文使用时序卷积网络模型，不仅兼顾了空间维度上的特征，而且还结合了时间维度上的特征，能够更加准确地检测 Fast-flux 僵尸网络流量。

3 实验过程及结果分析

3.1 实验环境

本文所使用的实验环境配置如表 1 与表 2 所示。

表 1 实验硬件环境参数

硬件	参数
内存	16GB
处理器	Intel(R) Core(TM) i5-8300H 2.30GHz
显卡	NVIDIA GeForce GTX 1050 Ti

表 2 实验软件环境参数

软件	参数
操作系统	Windows 10

编辑器	PyCharm
Python	3.8.12
PyTorch	1.10.1+cu113

3.2 测评方法

本文选择以下三种标准评价提出的方法：准确率(accuracy)、召回率(recall)和精确率(precision)。

准确率（Accuracy）计算公式为

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

召回率（Recall）计算公式为

$$Recall = \frac{TP}{TP + TN} \quad (2)$$

精确率（Precision）计算公式为

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (3)$$

为了确保训练阶段结果的公正性，数据集以7:3 的比例划分训练集与测试集。

3.3 方法参数选择

本文提出的检测方法需对原始 pcap 文件的数据进行过滤，再裁剪出 DNS 报文数据，然后统一化尺寸并转换成灰度图进行模型的训练。为了能够找到较为合适的数据大小，本文分别选取数据前324、361、……、625、676 字节，构成 18*18、19*19、……、25*25、26*26 的图片进行训练，各尺寸的图片训练结果如下图 12 所示。

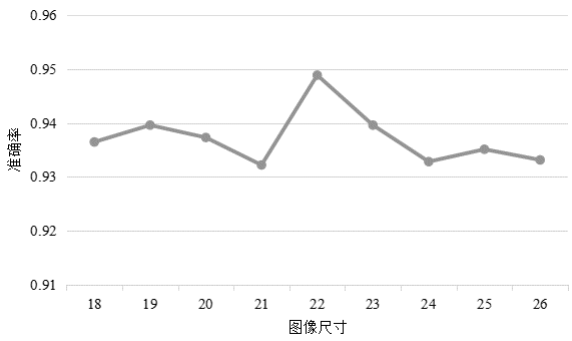


图 12 图片尺寸实验结果

通过图 12 可以看出，图像尺寸在 22 时为当前最优解，整体呈现出了正态分布的趋势。经过原始数据集的统计分析，发现原始数据集中超过84.5% 的数据大小是不超过 484 字节的。所以，在充分考虑模型训练效果的先决条件后，合理地减少样本中数据的大小，提高训练模型的效率。本文将流量数据经过预处理后转换为 22*22 的图像进行研究。

经过上述图像尺寸选择等一系列实验，最终本文模型指定超参数如表 3 所示。本文使用 ExponentialLR 方法进行调整学习率，其使用的参数 gamma 的值如下表所示。

表 3 模型超参数

字段	值	描述
levels	8	隐含层数
nhid	16	每个隐含层中包含的神经元数
image_size	22	输入图像的尺寸
ksize	4	卷积核大小
lr	1e−3	学习率
optim	'Adam'	优化器
gamma	0.78	学习率调整幅度
epochs	25	训练轮数
dropout	0.05	每个隐含层的神经元抛弃率

3.4 实验结果

使用本文设计的方法在经过 25 轮的训练后，准确率达到了 94.9%，召回率达到了 95.3%，精确率达到了 94.4%，如图 13 所示。从实验结果可以看出，本文提出的检测方法能够较好地检测 Fast-flux 僵尸网络，与传统的机器学习相比，也无需人工提取特征值。

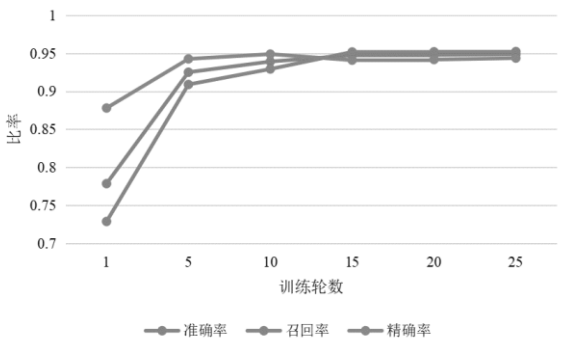


图 13 实验结果

4 结束语

随着僵尸网络隐蔽技术和控制手段的不断发展，传统的基于机器学习的 Fast-flux 僵尸网络检测方法越发难以有效检测，而且还存在着人工提取、处理特征等一系列繁琐困难的问题。为此，本文使用时序卷积网络模型进行检测 Fast-flux 僵尸网络，本方法在数据处理方面，仅保留 DNS 报文部分，减少数据运算量，提高训练速度。除此之外，不仅运用了空间维度上的特征，还兼顾了时间维度上的特征，避免了繁琐的数据处理，减少了人工参与选择的误差。实验证明，使用该模型的检测方法具有较高的准确率、精确率与召回率。

参考文献：

[1] 诸葛建伟,韩心慧,周勇林,叶志远,邹维.僵尸网络研究[J].软件学报,2008(03):702-715.

[2] 国家计算机网络应急技术处理协调中心.2020 年中国互联网网络安全报告[M].北京:人民邮电出版社,2021.

[3] Symantec.2019 年互联网安全威胁报告[EB/OL].https://docs.broadcom.com/docs/istr-24-2019-en. 2019.

[4] Fortinet.全球威胁态势报告[EB/OL]. https://www.fortinet.com/content/dam/fortinet/assets/threat- reports/threat-report-q4-2019.pdf.2019.

[5] 李可,方滨兴,崔翔,刘奇旭.僵尸网络发展研究[J].计算机研究与发展,2016,53(10): 2189-2206.

[6] 靳冲.Fast-Flux 网络检测与分析技术研究[D].北京:中国科学院研究生院.2011.

[7] Zang X D, Gong J, Mo S H, et al. Identifying Fast-flux Botnet With AGD Names at the Upper DNS Hierarchy[J]. IEEE Access, 2018, 6: 69713-69727.

[8] Guo X, Cheng G, Hu Y, et al. Progress in command and control server finding schemes of Botnet[C]//2016 IEEE Trustcom/BigDataSE/ISPA. IEEE, 2016:1723-1727.

[9] Guo Z, Guan Y. Active Probing-Based Schemes and Data Analytics for Investigating Malicious Fast-flux Web-Cloaking Based Domains[C]//2018 27th International Conference on Computer Communication and Networks (ICCCN). IEEE, 2018:1-9.

[10] Messabi K A, Aldwairi M, Yousif A A, et al. Malware detection using DNS records and domain name features[C]//Proceedings of the 2nd International Conference on Future Networks and Distributed Systems. ACM, 2018:29.

[11] Ruohonen J, Hyrynsalmi S, Mishkovski I, et al. The Black Mark Beside My Name Server: Exploring the Importance of Name Server IP Addresses in Malware DNS Graphs[C]//2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW).IEEE, 2016:264-269.

[12] Cafuta D, Sruk V, Dodig I. Fast-flux Botnet Detection Based on Traffic Response and Search Engines Credit Worthiness[J]. Tehnički vjesnik, 2018, 25(2):390-400.

[13] 张玉,刘纪伟.基于被动 DNS 流量的 Fast-Flux 域名检测方法[J/OL].南京邮电大学学报(自然科学版),2021(04):74-81[2022-03-18].DOI:10.14132/j.cnki.1673-5439.2021.04.010.

[14] 罗扶华,张爱新.基于深度学习的僵尸网络检测技术研究[J].通信技术,2020,53(01):174-179.

[15] 吴迪,方滨兴,崔翔,刘奇旭.BotCatcher:基于深度学习的僵尸网络检测系统[J].通信学报, 2018,39(08):18-28.

[16] 赵相男.基于模糊聚类的僵尸网络反规避技术研究[D].北京交通大学,2018.

[17] 周东杰.对域名系统新型扩展及安全问题的测量研究[D].战略支援部队信息工程大学,2019.DOI: 10.27188/d.cnki.gzjxu.2019.000085.

[18] University of Victoria.ISOT Botnet Dataset[EB/OL]. https://www.uvic.ca/ecs/ece/isot/datasets/botnet- ransomware/index.php.2010.

基于 5G 网络的危重症及突发公共卫生救治专网平台建设方案

竺智荣[1] 缪崇[1] 陈锦莹[1] 叶峰[1] 郑礼泷[2] 叶兴贵[2]
（1.福建省妇幼保健院，福建 福州 350001；
2.中国联通福建省分公司，福建 福州 350001）

摘 要：以快速响应和快速救治为出发点，依托高安全、低时延、大带宽的 5G 网络，结合医学急救、公共卫生等多学科联合，提出一种公共服务网络平台建设方案。引入 RAN 切片、边缘计算以及“风筝”技术，提高 5G 网络的数据安全与传输效率；通过多重校验与密码传输相叠加的方式，保障平台应用的保密性和完整性；结合政策法规与实际环境，借鉴最优秀的管理机制，构建适合于专网平台的安全管理体系。方案建设至今，已取得一定的社会效应和经济效应，并具备较高的安全性。

关键词：医学急救；公共卫生平台；5G 网络；密码技术

0 引言

“健康中国”与“互联网+”自十八大以来，被提升到了统筹全国整体战略层面，为贯彻全国健康医疗大数据的计划，福建省政府制定了《健康福建 2030”行动规划》和《福建省人民政府办公厅关于加快推进“互联网+医疗健康”发展的实施意见》(闽政办〔2018〕90 号)等文件，《福建省促进大数据产业发展行动计划(2018-2020 年)》是发展改革委印发的相关内容，为促进全省的工作落地。

从区域上看，我国大城市近年来医疗水平发展较快，但各地医疗水平却极不均衡，部分偏远地区较为落后，人民群众很难得到及时优质的医疗服务[2]。而一些远程会诊技术可以有效缓解各区之间的不平衡，例如通过高清晰的视频转播能力，在危机情况下实现异地治疗，利用互联网技术连接医院和医院、医院和病人，使得因交通、住宿和旅途等造成疾病延的可能性大大降低。但是，当前的医疗信息化还存在着诸多缺陷。远程视频等医疗场景由于受网络结构和医疗信息化设备的限制，容易出现延时卡顿以及数据泄露等问题。随着 5G 技术的横空出世与迅速发展，逐渐引起医疗界的广泛关注，依托于它的加强移动宽带(EMBB)、超可靠低时延通讯(URLLC)和海量机端连接(MMTC)三大技术优势，主要应用于低时延高可靠性场景、多热点高容量场景、连续广域覆盖场景以及低功耗大连接场景[1]。医疗领域信息化应用 5G 网络切片技术、边缘计算能够提升远程通信能力、降低移动设备能耗、优化医疗服务流程和促进智慧医院发展，使医疗信息化服务质量得到大幅提升。

福建省妇幼保健院担负着全省妇女儿童保健技术业务指导和妇女儿童常见病、多发病及疑难病症的诊治、教学、科研工作，需要进一步发挥自身学科优势。本方案依托福建省妇幼保健院医疗救治中心先行先试，根据医院建设要求与基础设施情况，以医院 5G 医疗专网为基础，融合网络安全、数据分析等技术，加强远程会诊、双向转诊、院前急救等平台建设，进一步发挥优质妇女及儿科优质医疗资源，为医联体、专科联盟成员单位提供远程协作和指导，将大大提高省内危重妇女儿童救治率和治愈率，积极探索突发公共卫生事件的应急处理办法和转运途中特定场景下的应对策略，降低出生缺陷和孕产妇死亡率。

1 方案建设目标和主要任务

1.1 建设目标

发挥福建省妇幼保健院妇女儿童专家资源优势，探索构建危重症及公共卫生救治新举措，打造安全高效的区域妇幼危重症急救与突发公共卫生事件应急协作网络平台，提高区域医疗整体救治水平、提升区域居民幸福指数，造福区域人民。

1.2 主要任务

本方案的主要建设任务包括：

(1) 面向危重症及突发公共卫生事件的 5G 专网，具体包括：

① 接入侧；

② MEC 业务系统；

③ 数据交互侧。

(2) 面向危重症及突发公共卫生事件的云上妇幼平台，具体包括：

① 移动办公；

② 指挥调度；

③ 院前急救；

④ 远程会诊；

⑤ 远程示教。

2 方案建设思路

2.1 总体架构

结合 2.2 所列主要任务，将建设一张福建省妇幼保健院 5G 医疗专网，并在此基础上，搭建危重症及突发公共卫生事件救治网络平台，用于突发公共卫生事件指挥调度、院前急救、医联体/专科联盟成员单位间远程会诊、远程查房、远程示教等远程医疗活动，同时满足医务人员移动办公等远程办公行为。因此，可进一步将所建设的 5G 专网分为三个子专网，即 5G 远程办公专网、5G 远程急救专网、5G 远程会诊专网。网络总体架构如图 3-1 所示，建设内容也将围绕这三张子专网分别构建安全可靠的福建省妇幼保健院 5G 医疗应用平台。

在这三张子专网建设的基础之上，将形成以福建省妇幼保健院为省级会诊中心，辐射福州、莆田、南平、三明、龙岩、宁德、厦门、漳州、泉州 9 地市共 15 个分中心，分中心下辖设县乡级 60 个协作点，形成省、市、县三级危重症及突发公共卫生救治的 5G 专网。本方案总体功能架构如图 3-2 所示。

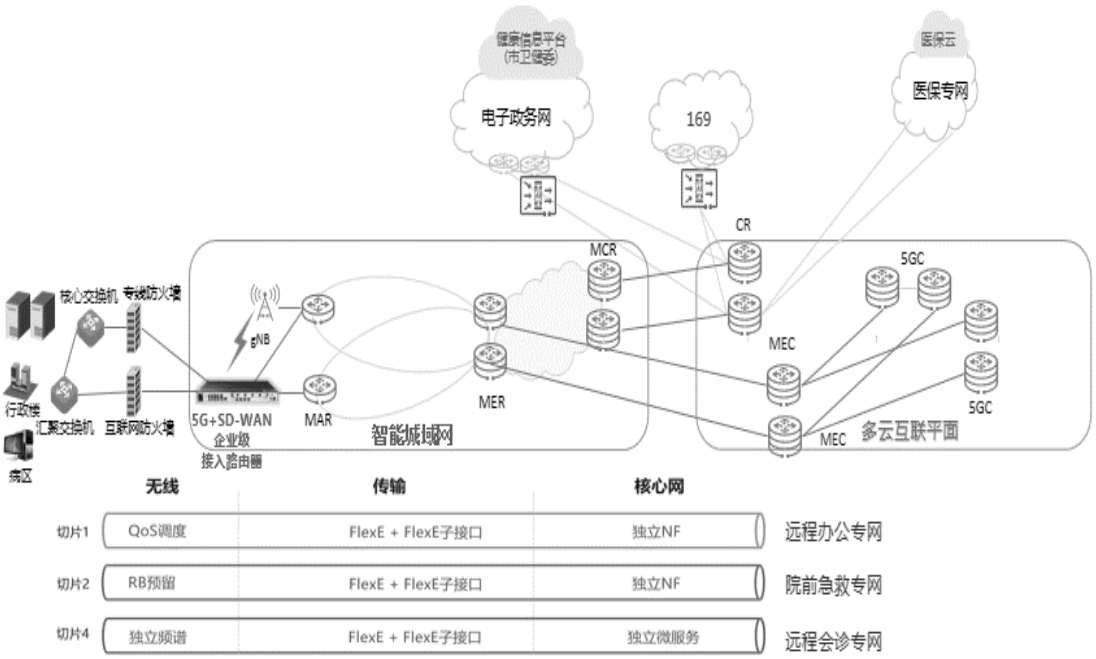


图 3-1 福建省妇幼保健院 5G 医疗专网总体架构

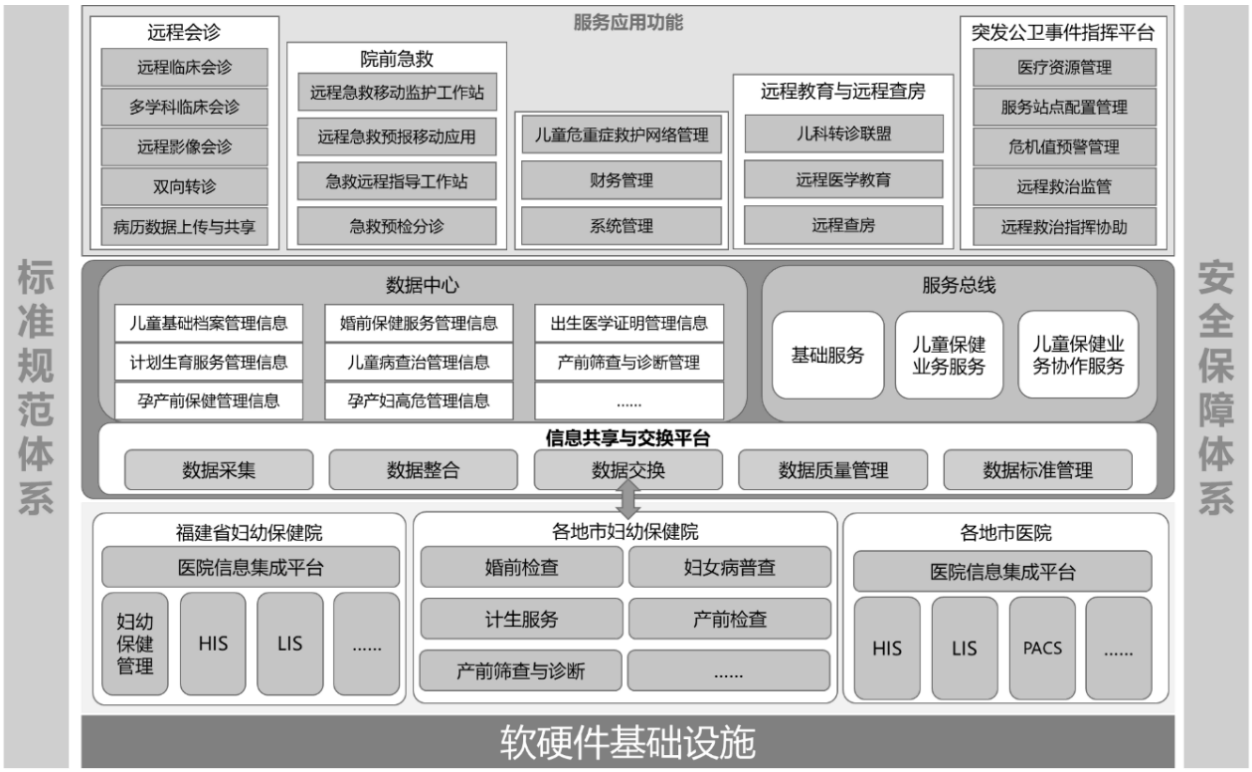


图 3–2 危重症及突发公共卫生事件救治网络平台功能架构图

2.2 技术路线

本方案拟采用的技术路线为：通过建设安全高效的 5G 专网，为面向危重症及突发公共卫生事件的网络平台提供安全的信息通路，通过网络安全、接入安全、系统安全、软件安全和管理安全等措施实现项目整体安全闭环管理。其流程图如图 3–3 所示。

图 3–3 技术路线

2.3 主要指标

专网应满足不同场景下的性能指标，包括但不限于重症监护、移动医护、视频会诊和移动办公，具体参数见表1–表4所示。

表 1 重症监护类性能指标

代表性场景	急危重症监护与会诊等（特点：多元数据、复杂数据集成）	
典型数据	4K 视频，图像（GB 级），体征数据	
网络技术配置要求（单场景单设备）	上行速率	>20Mbps
	下行速率	>5Mbps
	网络时延	<100ms
	可靠性要求	99.999%
	网络抖动要求	<20ms

表 2 移动医护终端类性能指标

代表性场景	医护查房，移动护理等	
典型数据	图像（GB 级），病历数据	
网络技术配置要求（单场景单设备）	上行速率	>2Mbps
	下行速率	>20Mbps
	网络时延	<100ms
	可靠性要求	99.999%
	网络抖动要求	<20ms

表 3 视频交互会诊类性能指标

代表性场景	远程会诊，床旁会诊，多学科会诊等	
典型数据	4K 视频，图像（GB 级）	
网络技术配置要求（单场景单设备）	上行速率	>20Mbps
	下行速率	>20Mbps
	网络时延	<100ms
	可靠性要求	99.999%
	网络抖动要求	<20ms

表 4 移动办公类性能指标

代表性场景	在线查询检查检验报告、病例数据，远程处理办公事务	
典型数据	图像（GB 级），病历数据，办公数据	
网络技术配置要求（单场景单设备）	上行速率	>20Mbps
	下行速率	>5Mbps
	网络时延	<100ms
	可靠性要求	99.999%
	网络抖动要求	<20ms

安全指标重点关注端到端的安全接入和信息传输保障，主要包括：

(1) 基础设施安全。应支持物理安全保护机制（如：防拆、防盗、防恶意断电、防篡改等，设备断电/重启、链路网口断开等问题发生后应触发告警）；应支持为硬件 WAN 口、LAN 口、串口等进行安全访问控制；应支持内置的安全功能；使用虚拟机或容器部署 UPF、MEP 以及 MEC APP 等，应支持资源的安全隔离、镜像和镜像仓库的完整性和机密性保护等；支持 MEC 节点级容灾满足可靠性的要求。

(2) 医院应用系统安全。应支持对访问进行认证和授权；应支持敏感数据安全保护，防止非授权访问、篡改等。

(3) 安全管理。应支持对其使用的操作系统、中间件、数据库以及 WEB 管理接口进行安全加固，满足安全基线的要求；应支持使用标准格式的证书，支持证书有效期管理、证书失效前预警；应支持流量过载控制；应支持使用安全工具对 MEC 平台进行扫描，保证不存在高危漏洞以及未使用、不必要的端口和服务等。

2.4 接入方式

5G 医疗专网访问，利用 5G AKA 和 EAP−AKA′ 鉴权算法，实现对于 5G 终端和 5G 网络的双向鉴权，保证特定的 5G 专网用户访问网络，并使专网用户和公网用户隔离起来。在 5G 终端和 5G 网络之间，实现基于 256 bit 的加密算法，确保数据的保密传输。

在 5G 安全机制基础之上，利用 VPN 机制或者终端识别机制，实现 5G 终端和医院内网服务器之

间的端到端的安全访问，提高终端访问的安全性。

方案在部署上相对于 4G 网络在物理位置、业务类型、网络架构等方面均发生了变化，使得传统网络架构和新架构存在巨大差异，需要做到如下安全考虑：

(1) 院区场景接入安全：院区设备不由院区控制和管理，院区希望非法用户不进院区，而且数据不出院区；

(2) 网络层接入安全：网络协议接口增多，攻击面增大，部署过程中存在边缘与中心互相攻击的风险；

(3) 终端接入安全：身份仿冒、信号欺骗与设备劫持等一系列安全问题。

针对安全需求设计接入安全方案如下图所示：

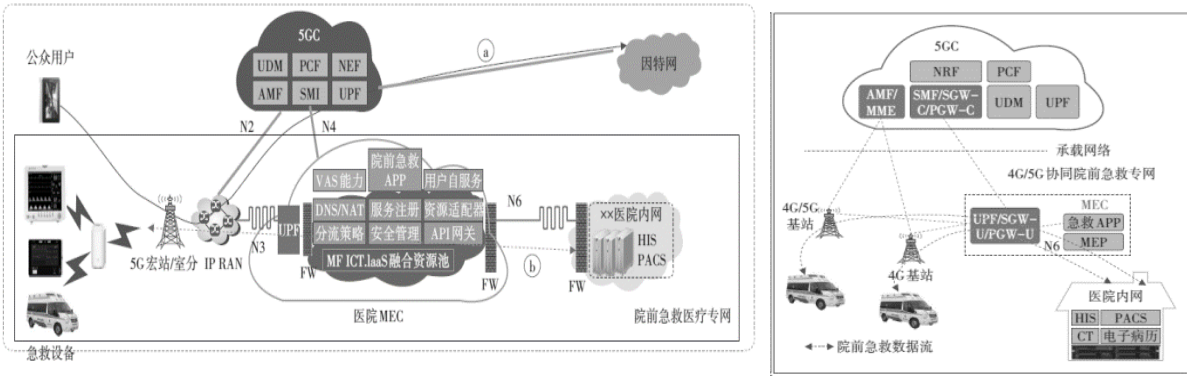


图 3–4 4G 接入 5G 升级方案

主要步骤包括：(1) AMF 升级支持 AMF/MME；(2) SMF 升级支持 SMF/SGW−C/PGW−C；(3) UPF（MEC 的转发网元）升级支持 UPF/SGW−U/PGW−U。升级之后，当用户设备接入 4G 网络时，通过识别用户，根据专用 APN/DNN 将当前业务分配至院区边缘的 UPF/SGW−U/PGW−U 承载，因此，保证了业务传输路由和通过 5G 接入时的路由一致。

3 方案建设内容

本方案采用统一规划、充分论证、分步实施的原则开展建设。建设内容主要包括：危重症与突发公共卫生事件 5G 专网和云上妇幼平台。以下进行具体实施方案说明。

3.1 面向危重症及突发公共卫生事件的 5G 专网

面向危重症及突发公共卫生事件的 5G 专网建设主要包括 5G 远程办公专网、5G 远程急救专网、5G 远程会诊专网三张子专网的建设。三张子专网的建设按照地域可划分为院区内与院区外。院区外子专网的建设主要利用运营商建立的无线接入网、承载网、核心网和数据交换网等通信大网，通过路由策略、控制转发等过程发送至院区内边缘 UPF。院区内专网的建设主要分为三大部分：(1) 接入侧；(2) MEC 业务系统；(3) 数据交互侧。

(1) 接入侧

接入侧主要是运营商利用 SDN 和 NFV 技术进行网络切片，在有限物理网络设备建立多种相互隔离的虚拟网络。根据各子网的性能指标与安全指标的不同，提供不同的通信方式与安全策略。接入侧的网络切片技术主要有基于 VPN+QoS、Flex−E 和以太独立端口的切片技术。

在搭建专网时，基于不同的隔离等级（如物理隔离、逻辑隔离）实现切片定制化构建。可利用专网卡、CPE 设备等方式进行管理控制。具体网络架构如图 4−1 所示。

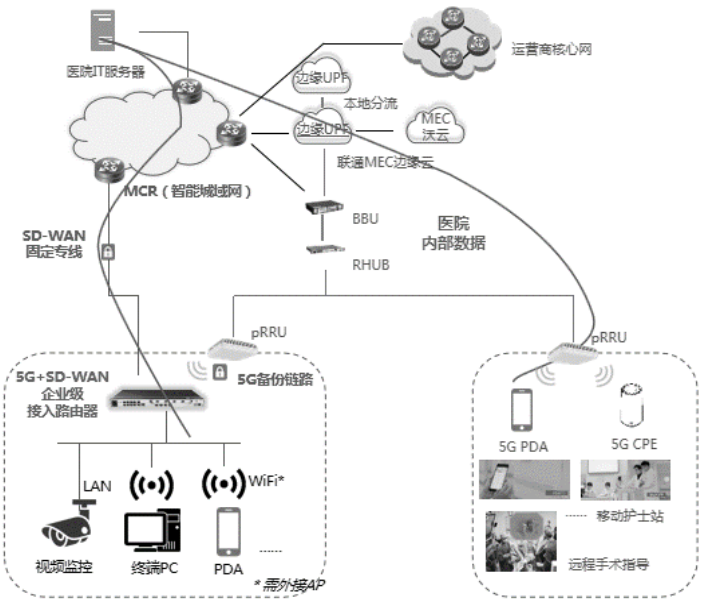


图 4–1 接入侧网络架构图

(2) MEC 业务系统

边缘业务平台架构包括硬件资源层、虚拟化层和平台能力层，在多接入的基础网络之上，为应用开发者提供灵活的平台能力和丰富的 API，赋能各行业应用。通过在医疗系统部署 MEC，可以释放医院网络压力，提升医院工作效率，帮扶基层医疗人员和机构，提高医疗服务水平和医院信息数据的安全。

方案拟在医院内新建一个 MEC 节点，MEC 平台与中国联通边缘业务管理平台进行对接，由 MEPM 平台进行管理和编排。

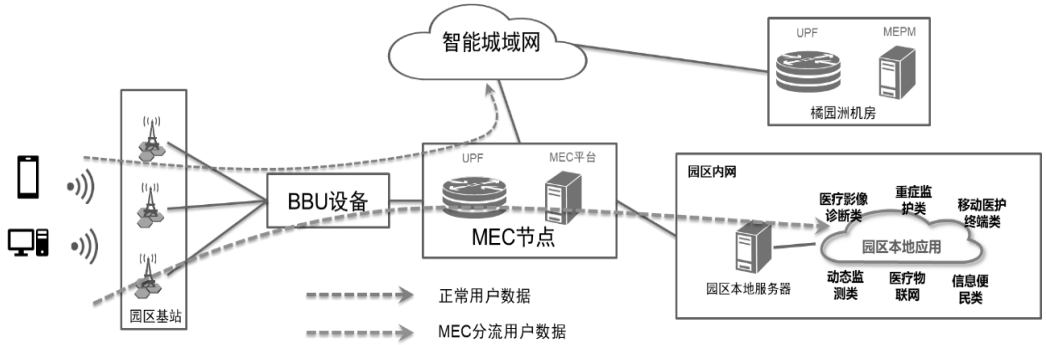


图 4–2 MEC 院区数据流向

MEC 院区数据流向如图 4–2 所示，基站上行数据经过 MEC 节点分流，正常用户数据直接进入运营商网络，院区本地数据分流进入院区内网，实现业务数据不出医院，满足用户时延要求。

MEC 采用安全策略主要包括管理平面（Underlay）和用户平面（Overlay）的防护。管理平面复用大网或安全建设能力，用户平面通过全国统一的 MEC 边缘云能力实现。

管理平面安全策略包括：云边协同中外部攻击防御、安全域隔离、内部子域隔离、边缘与核心隔离、物理环境安全、APP 应用内部安全、管理网络和用户平面的隔离等。

用户平面安全策略包括：APP 应用内部安全、租户应用隔离安全、云边协同出口安全、应用访问

鉴权和熔断等。

(3) 数据交互侧

业务流向共有 3 个，分别为医院内业务与 MEC、MEC 与医疗数据中心、基站与 5GC 控制面流量。

① 医院内业务与 MEC

该流向业务需承载网支持 FLEX-E 切片，根据 MEC 部署地理位置的不同，承载网分为两大类，第一类模型为 MEC 部署在医院院区；第二类模型为 MEC 集中部署在区县汇聚节点。

对于院内部署 MEC，医院内网业务和 5G 基站 DU 通过 MEC 节点进行数据回传。

对 MEC 集中部署在区县汇聚节点，医院内网业务和 5G 基站 DU 通过院区内部光缆均连接至医院院区新建的 IPRAN2.0 设备，设备与区县 MEC 节点新建的 IPRAN2.0 设备组网。5G 基站的控制面流量通过移动回传网传送至 5GC，其他流量通过移动回传网对接至区县 MEC 节点。

② MEC-省医疗数据中心

该流向业务采用省内 PeOTN 网络进行传送，PeOTN 网络可提供时延最低且为硬管道隔离的承载链路。需要在 MEC 节点新建 PeOTN MCE 设备，新建组网光缆。

③ 基站-5G 控制面流量

通过移动回传网至 5GC，可利旧现有网络，再通过联通独家的“风筝”技术，在边缘 UPF 部署应急控制面，当大区控制面中断，应急控制面接管业务，网断业务不断，实现高可靠性。

3.2 面向危重症及突发公共卫生事件的云上妇幼平台

面向危重症及突发公共卫生事件的云上妇幼平台建设内容主要包括：移动办公、指挥调度、院前急救、远程会诊和远程示教等。

(1) 移动办公

移动办公是移动通信网络的重要应用之一，在日常医疗办公应用中，主要利用无线接入手段，实现区域的完全覆盖，信息上传下达全过程互动与交流，建立多部门协同体系，大大提升办事效率。为实现移动办公的数据安全传输，采用 5G 端到端解决方案，其架构图如图 4-3 所示。

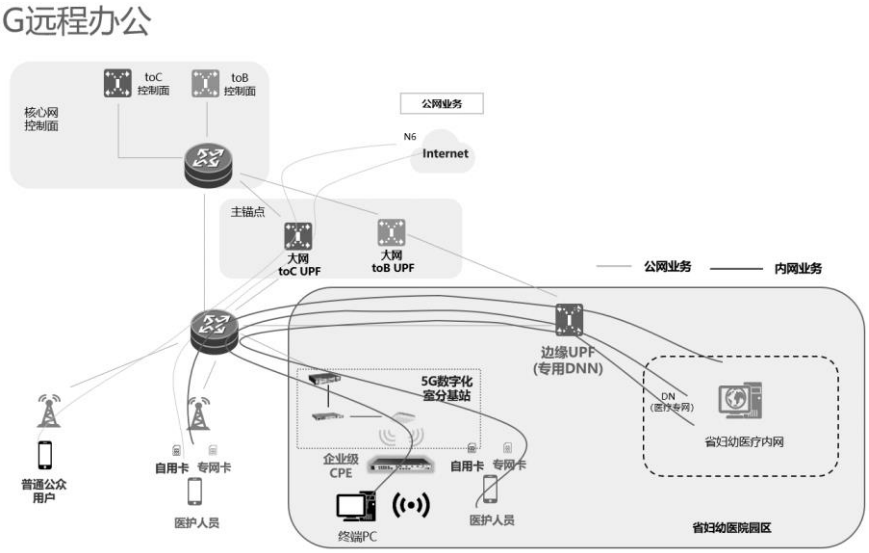


图 4-3 5G 远程办公专网

在 5G 远程办公专网中，医护人员通过 5G 专用物联网卡实现省内任意地点安全访问院内系统，医护人员的手机信号通过 DNN 信息后将数据包发往指定的医院边缘 UPF，实现全程全网安全通道访问办公网，随时随地查看病人实时病历、进行公文处理等移动办公、移动诊疗操作。

在医院院区范围内，利用运营商靠近基站边缘的 MEC 部署方式，将 MEC 节点部署于基站与运营商核心网之间，通过业务分流将数据发送至院内相关网络设备。通过该方式，无需经过运营商核心

网，一方面能够降低传输时延，提升网络响应效率和用户体验。另一方面通过业务控制流与用户数据流的分离，避免内部数据泄露。

而在院区范围外，运营商根据用户专属卡识别用户身份，利用运营商大网 UPF 实现业务数据的路由和转发、数据和业务识别、动作和策略执行等操作，最终将数据转发至院区的边缘 UPF 设备上进行下一步的数据处理。

(2) 指挥调度

建立一套运转高效、科学规范的公共卫生指挥决策平台，可大大提高区域卫生应急指挥和应急处置的科学化、信息化与智能化水平，有效提升突发公共卫生事件的处理能力[5]。基于公共卫生监测数据收集、分析，实施突发公共卫生事件相关信息监测预警，有助于实现突发事件的早期预测预警，及时采取科学、有效应对措施，尽可能地避免或降低事件对民众身体健康和生命安全所造成的危害[6]。

本方案将建设如图 4−4 的统一调度平台，实现医疗资源的充分调度。利于 5G 无线网络灵活组网，超大带宽和不受传输距离限制等特点，医护人员可将现场情况实时回传至指挥平台，便于专家远程指挥救助。平台支持位置共享，可以随时精准地展开救治工作。在发现紧急情况后，紧急救助人员可第一时间获取信息，并通过集群对讲调度医疗队赶赴现场，对突发疾病现场进行紧急救治，同时将现场视频回传至调度室，以便进行远程指挥调度。

(3) 院前急救

正确运用院前急救措施，对我院呵护产妇及胎儿生命健康、生存质量都具有积极的作用。此外，由于区域内医院间新生儿救治水平参差不齐，有必要建立区域性危重症新生儿转诊网络[7]。院前急救平台在急危重症孕产妇院前急救与转运、新生儿转诊过程中扮演着重要的角色[8]。

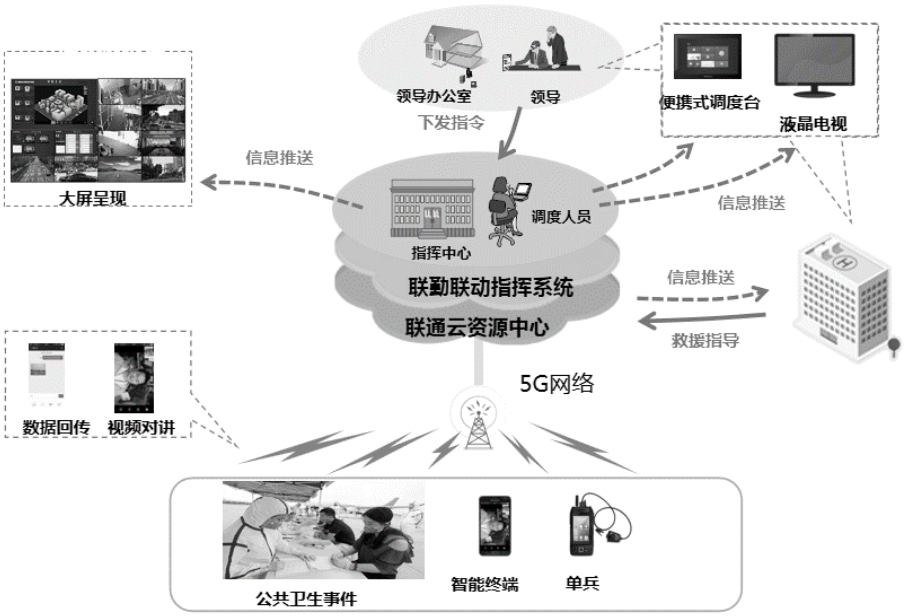


图 4−4 公共卫生事件指挥调度平台

院前急救平台利用 5G 网络实现远程医疗监护与监测，对患者的生命体征进行实时、持续和长时间的监测，并将医疗设备获得的生命体征数据和危急告警信息从院前的救护车上传送给院内急诊人员，实现院前急救与院内急诊的全流程无缝连接。基于 5G 网络实现车辆定位等信息的实时传输，实现智能车辆管理与智能调度，同时传输现场视频，实现远程实时会诊与指导。使用机器学习算法或产品，对收集的监测数据进行模型分析，发现异常情况及时报警，并针对特殊病种的影像学结果采用智能分析，同时为医生提供辅助决策。具体网络架构如图 4−5 所示。

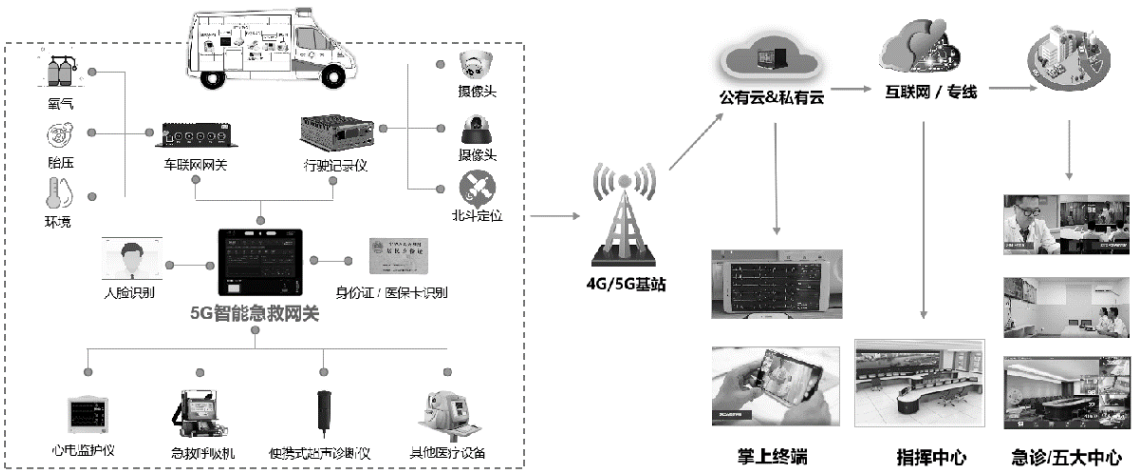


图 4–5 5G 院前急救平台网络架构

该平台整个数据传输过程基于院内集成平台和标准的数据传输格式，以保证院内系统间的互联互通。系统功能架构如图 4–6 所示，主要包含以下四大功能模块：① 院前急救业务模块。主要负责将院前急救患者的生命体征、病程记录等数据传输到院内。② 院内急救业务模块。主要包含挂号系统、医嘱系统、电子病历系统、医技系统等急诊临床业务子系统。业务数据封装后，通过消息队列的方式和院内临床数据中心进行数据交互。③ 院前院内衔接模块。这是一个数据交换模块，通过标准的 Web Serverice 接口实现院外数据到院内数据的传输，同时利用网络安全设备保证传输安全，利用数据签名与加解密技术保证数据有效性与完整性。④ 基础设施模块。在系统运行和互联互通过程中，使用 RFID 技术、5G 高效传输技术、自然语言处理技术处理相关数据等。

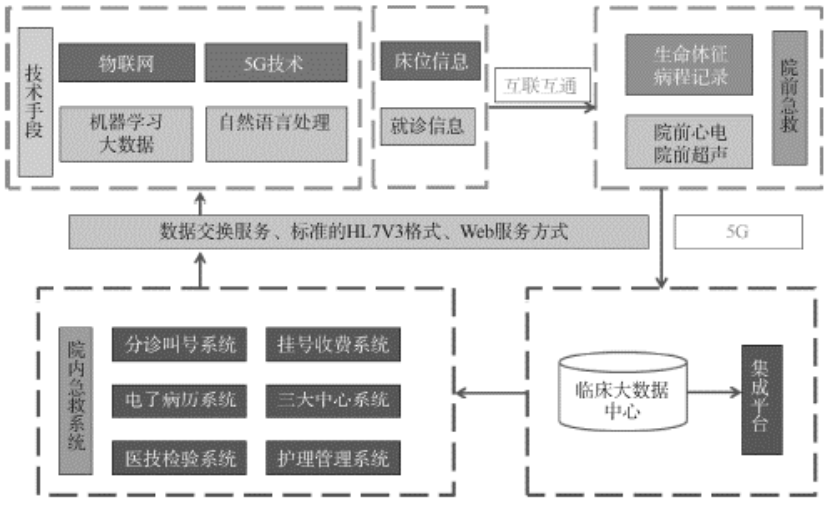


图 4–6 基于 5G 的院前急救平台功能架构图

(4) 远程会诊

基于 5G 的远程临床会诊平台，利用 5G 无线网络、VPN 等技术，搭建满足不同院区、不同地理位置的远程临床会诊需求。利用通信技术手段共同探讨患者病情，进一步完善并制定更具针对性的诊疗方案[9]。在福建省妇幼保健院、福建省妇产医院、福建省儿童医院三院间的两两院区通过 2 根光纤互连，实现一主两备，高效协同的网络架构。与医联体单位、专科联盟单位的远程服务，通过运营商 5G 网络建立 IPSec VPN 通道，利用边界防火墙、

边界网关等网络设备保证网络传输安全与稳定。具体部署方案如图 4-7 所示，VPN 通道部署方案如图 4-8 所示。

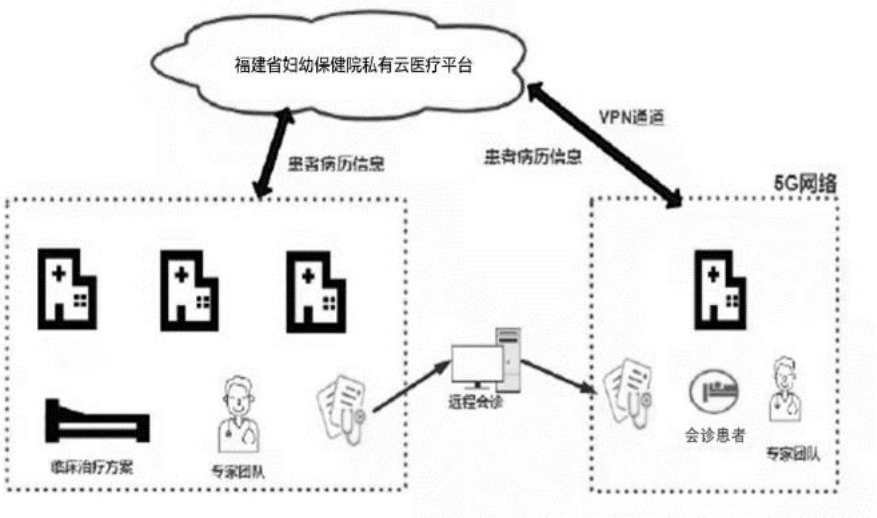


图 4-7 5G 远程临床会诊平台网络总体部署架构图

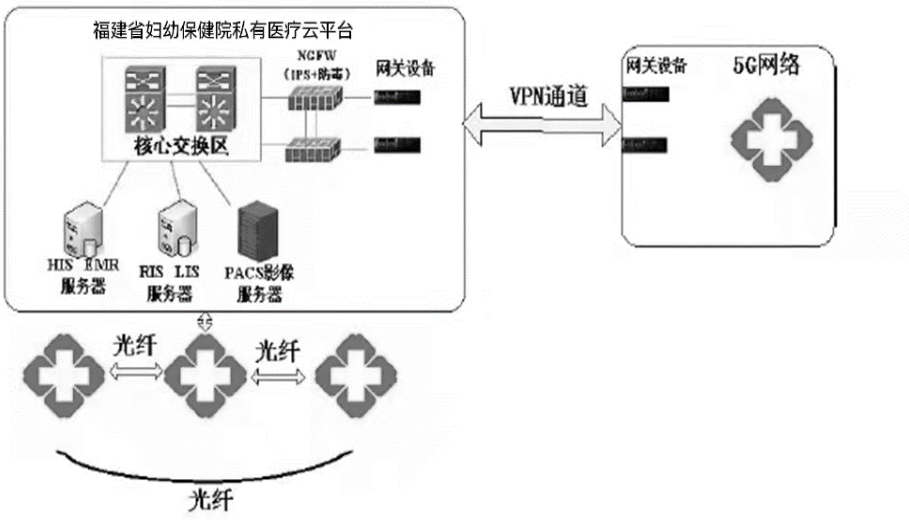


图 4-8 VPN 通道部署方案

(5) 5G 示教平台

医疗教育指面向医疗卫生技术人员进行的教育培训，用户包括医疗、护理、医技人员。远程医学教育培训主要包括：基于音视频会议系统的教学平台、基于使用场景的教学平台和基于 VR/AR 设备的教学平台三类产品形态。

以远程手术示教为例，通过在医院部署远程医疗视讯协同终端，配合融合 5G 医疗专网，赋能上级医院手术室、示教室与医联体、联盟单位手术室、学习室之间建立全景协作通路，实现上级医院、多医联体、联盟单位之间的多点、多向、实时、高清、随时、随地的音视频和手术全景图像传送。实现手术示教实时同步直播给医联体、联盟单位，学习更直观，且可以实时视频互动交流，直播结束后可回放，反复观看[10]。该平台网络架构如图 4-9 所示。

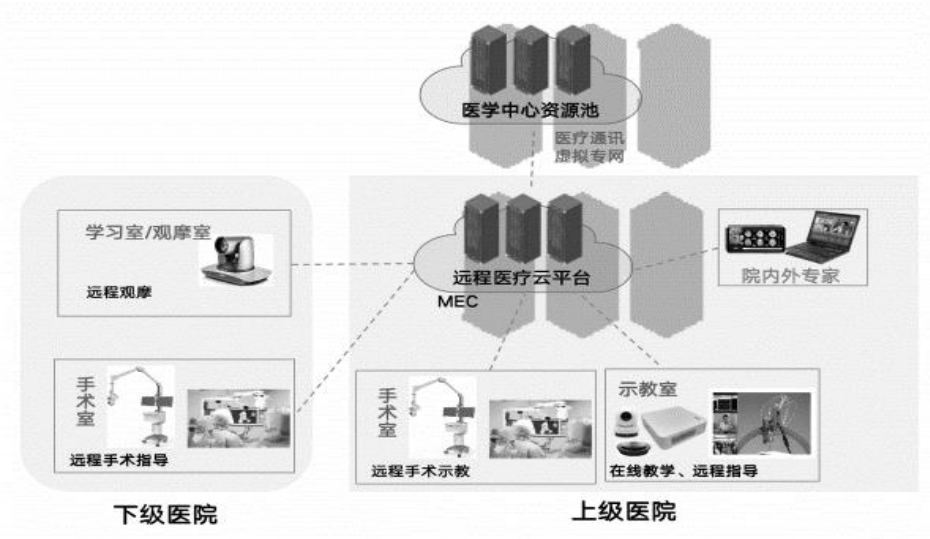


图 4–9 5G 手术示教平台网络架构图

4 创新性

4.1 专网安全创新性

5G 专网技术在基础网络的传输性能、安全性能以及便利性能上，较传统的网络基础方案有明显的提升。随大网演进，技术迭代永不落伍，并可提供端到端精细规划、设计、建设、维护及优化服务，实现覆盖、速率、容量、上下行配比的灵活配置。

(1) 采用 RAN 切片技术以及 RB 资源预留技术，从频谱到核心网（边缘 UPF），实现端到端切片，医疗业务完全与公网隔离，构建超高安全、超高可靠 5G 专网

(2) 入驻式边缘 UPF+应急控制面(风筝)，当大区控制面中断，应急控制面接管业务，保证网断业务不断；

(3) 边缘 UPF 入驻部署至院内机房，实现数据不出医院；

(4) 为医护人员配置专属 DNN 卡号，与一般用户实现高安全隔离；

(5) 轻量级部署，快速极简；随大网演进，技术迭代永不落伍。

4.2 平台安全创新性

(1) 提供一种在会话级别安全验证用户的方法，并且在系统最初建立身份时或者有迹象表明身份可能已被泄露时，利用其他方法或技术进一步验证用户的身份；

(2) 使用管理手段、物理方法和技术保护措施来保护数据信息免遭未经授权的泄露或访问；

(3) 远程访问或特权访问要求双因素身份验证以降低未经授权访问的风险；

(4) 采用校验技术与密码技术保证医疗数据在传输过程中的保密性、完整性。

4.3 安全管理创新性

本方案同步制定《危重症及突发公共卫生事件安全管理制度》，邀请行业内权威专家对安全制度进行评审，定期对安全管理制度进行动态维护，根据国家及行业法律法规结合医院业务需要及时对安全管理制度进行更新。同时，本方案成立由院领导挂帅的安全管理委员会，对涉及的各安全环节设置工作岗位，明确岗位职责，对安全环节进行闭环管理考核。

5 项目运行情况

医院内网通过部署应用安全设备与安全技术，在 2021 年福建省卫健委组织的网络安全攻防演练中，定位到明确的外部攻击源 20 多个，提交了防守方报告，最终排名并列省属医院第三。此外，成功拦截不明攻击源不计其数，攻击形式主要是以端口恶意扫描、弱口令漏洞、数据 SQL 注入漏洞、未授权访问漏洞、远程代码执行等漏洞进行网络攻

击。

成功开展多次远程会诊和远程示教，充分发挥信息化在分级诊疗中的支撑作用，有效畅通“优质妇儿医疗资源”下沉渠道，积极带动下级医院专科诊疗水平和服务能力的提升，努力为全省人民群众提供“同质化、高水平”的诊疗服务。

图 5-1 福建省妇幼保健院专家远程指导超声检查

6 结束语

近年来，远程会诊和互联网医院正逐步建立服务体系、保障支撑体系和管理体系，以改善优质医疗资源区域性不平衡、基层临床诊疗能力差和医患互信度低等问题。本方案以负责业务建设的福建省妇幼保健院牵头危重症及突发公共卫生救治平台建设，福建联通负责项目涉及的 5G 专网和云平台建设，确保专网安全、患者信息安全和医院数据安全。方案建设至今，已取得一定的社会效应和经济效应，并具备较高的安全性。

参考文献

[1] Minahil, Ayub M F, Mahmood K, et al. Lightweight authentication protocol for e-health clouds in IoT based applications through 5G technology[J]. Digital Communications and Networks, 2020.

[2] Duan W, Ji Y, Zhang Y, et al. 5G Technologies Based Remote E-Health: Architecture, Applications, and Solutions[J]. 2020.

[3] Ybs A, Jb A, Ap B, et al. Quality of perception prediction in 5G slices for e-Health services using user-perceived QoS[J]. Computer Communications, 2021.

[4] Ahad A, Tahir M, Sheikh M, et al. A Game Theory Based Clustering Scheme (GCS) for 5G-based Smart Healthcare[C]// 2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT). IEEE, 2020.

[5] Hussien N S, Khafidz A I, Masmuzidin M Z. The Enhancement of First Aid Treatment for Medical Facilities[J]. Journal of Physics Conference Series, 2020, 1529:052098.

[6] Yin S, Liu J, Teng L. Improved Elliptic Curve Cryptography with Homomorphic Encryption for Medical Image Encryption[J]. International Journal of Network Security, 2020, 22(3):421-426.

[7] 乔莉, 张劲松. 5G 对急救体系的影响及研究现状[J]. 中华急诊医学杂志, 2020, 29(7):3.

[8] 张夏童, 任智源, 胡锦涛,等. 面向医疗大数据任务低时延需求的路径计算方案[J]. 西安交通大学学报, 2020, 54(2):8.

[9] Alrazgan M. Internet of Medical Things and Edge Computing for Improving Healthcare in Smart Cities[J]. Mathematical Problems in Engineering, 2022, 2022.

[10] Mobile Edge Computing Enabled 5G Health Monitoring for Internet of Medical Things: A Decentralized Game Theoretic Approach[J]. IEEE Journal on Selected Areas in Communications, 2021, 39(2):463-478.

[11] Ranaweera P, Liyanage M, Jurcut A D. Novel MEC Based Approaches for Smart Hospitals to Combat COVID-19 Pandemic[J]. IEEE Consumer Electronics Magazine, 2021(2).

[12] Braeken A, Liyanage M. Highly Efficient Key Agreement for Remote Patient Monitoring in MEC enabled 5G Networks[J]. The Journal of Supercomputing, 2020(5).

政企事业单位采购业务相关数据安全治理解决方案

李 荏
（山石网科通信技术股份有限公司，江苏 苏州 215000）

摘 要：政企事业单位采购业务涉及数据具有体量大、类型杂、价值高等特点，数据安全治理的开展迫在眉睫。通过组织建设、现状摸底、数据分类分级、风险评估、体系建设、持续面向全员的培训等步骤可为政企事业单位提供一套切合自身业务的认知理论和关键落地路径，提升采购业务领域保密工作整体防护能力和效率，杜绝信息泄密事件，防控廉洁风险。

关键词：数据安全治理体系；数据分类分级；基于语义特征的内容识别

0 前言

数字化正成为引领中国经济高质量发展的新引擎[1]。在 5G、云计算、物联网、大数据、移动互联、区块链等新兴科技的加持下，我国不同行业均迎来了一场数字化转型升级的变革。

政企事业单位的数字化发展对其网络安全、信息安全提出了更高的要求。远程办公、内外业务协同、组织分支互联等需求的变化，让事业单位边界逐渐泛化，极大的增加了其安全防护风险。过往粗放型的数据使用和管理体系已很难适应当下的数字化业务系统。如何防护敏感数据泄露，让数字化建设与网络安全保障双强化，成为政企事业单位迫在眉睫的问题[2]。另一方面，伴随着信息安全形势的变化和发展，数据的重要性越来越凸显，相关的法律法规、标准要求、行业指南等层出不穷，不同的监管机构均对如何保护组织和个人的敏感数据提出了对应的监管要求，从趋势来看监管机构对于数据保护的要求越来越细化，而且监管也越来越严格和具备强制性[3]。因此，如何满足不同监管机构的数据安全保护要求已经成了合规的重要议题。

在政企事业单位众多数据安全治理场景中采购业务便是一个难以绕开的话题。政企事业单位每年都有近百亿资金需对外采购，廉洁保密一直是工作重点。采购业务泄密风险防控工作面临“点多、面广、战线长、参与人员多”的情况，且涉密信息受到投标人及利益相关方的高度关注，采购业务面临内外部利益诉求交织，泄密事件成为了“围猎”与“被围猎”之间的纽带，政企事业单位的信息安全和廉洁从业安全面临非常严峻的挑战。仅依靠相关人员安全意识的提升和制度的完善，难以做到泄密事件的事前预防和事后追溯，必须依靠相应的技术和管理手段，对采购业务领域涉密资料的全过程监督和管控，才能有效开展保密工作。管理者亟需通过相关措施来建立组织内部整体数据安全治理体系，有效支撑采购业务的核心资料的保密工作。

1 政企事业单位采购业务相关数据安全治理解决方案

相对于传统网络安全规划建设，数据安全治理对“科学性”、“系统性”提出更高要求，而由于数据与业务相伴相生，关系密切，所以政企事业单位采购业务相关数据安全治理又必须通业务场景与数据流转深度绑定[4]。通过组织建设、现状摸底、数据分类分级、风险评估、体系建设、持续面向全员的培训等步骤有序、科学地完成政企事业单位采购业务相关数据安全治理建设工作。

1.1 组织建设

无论是数据安全法还是关基条例，都要求数据安全治理要有专门的组织和人员负责，并定岗定责。在政企事业单位采购业务相关数据安全治理的过程中，成立专门的数据安全治理机构是首要条件[5]。该组织结合政企事业单位的战略方针，制定符合自身的数据安全治理的政策，并落实和监督政策有效执行。

1.2 现状摸底

为了对政企事业单位采购业务相关数据进行综合安全治理，首要任务便是对数据资产现状进行清查摸底。通过多种方式来发现数据所有者、存储位置、整体业务架构等信息。因此首先需要对政企事业单位的用户进行前期摸底调研：了解了组织关于数据安全治理的战略方针，组织有哪些业务及相应的业务系统，业务系统开放形式，访问方式、交互方式，业务系统相关数据是否涉及个人隐私信息，数据存储的形式、位置及访问的方式，数据的重要性、影响范围和影响程度，现有信息系统建设及数据安全建设情况等等；同时目还需参考《网络安全法》、《数据安全法》、《个人信息保护法》等法律中的相关规定要求，合法合规管控涉密资料的流转。

1.3 数据分类分级

基于现行行业标准与安全实践经验制定合理、有效的数据分类分级规则，对政企事业单位采购业务相关数据资产进行全盘梳理，为实现“核心数据安全优先，其余效率优先”的差异化防护打下基础。通过访谈方式、文件审核、查看系统、查看数据库等方式，梳理现有政企事业单位数据覆盖的数据范围及预测未来可能覆盖的数据范围，形成数据目录结构；依据国家相关的法律法规及行业规范，考虑数据对国家安全、社会稳定和公民安全的重要程度，结合以往项目实践经验，完成对政企事业单位的数据分类分级指导规范；依据数据分类分级指南，通过分类分级工具结合人工的方式，对政企事业单位的数据进行标签管理，明确数据的类别级别；同时随着政策变化和日常实际运营情况，及时对数据的分类分级指南、工具内策略及时调整更新。

具体来看，针对结构化数据通过部署数据库扫描监控系统实现对敏感数据自动发现、分析和梳理，将敏感数据按照内置或自定义策略进行分类分级，确保政企事业单位用户对大数据平台内的敏感数据透视，为用户对敏感数据的使用和管理提供依据。同时通过风险扫描和监控功能确保数据库漏洞和不稳定因素提前发现提前处置，确保数据库安全稳定运行。

针对非结构化数据采用基于语义特征的自然语言处理内容识别技术对政企事业单位采购业务相关数据大规模实时精准分类。该技术通过无监督机器学习引擎来分析大量未经标注的原始文档集，自动按照内容进行主题梳理，并通过人工干预灵活调整语义相似度，获得满意的聚类效果。将聚类结果作为标注样本，实施有监督机器学习，提取短句或长组合词作为语义特征，自动生成分类规则库。在此过程中，用户亦可人工干预特征选择，进行漏报及误报样本的提取及规则优化，使用反向对照样本加强训练，进而提升工作效率及分类准确性。将文本分类规则推送至部署在组织中端点、服务器、和网络等处的轻量化分布式分类器，即可实时感知关键数据的分布和使用状况，为数据安全治理提供基础支撑。

1.4 风险评估

完成分级分类过程后，结合风险评估和数据安全能力成熟度模型发现政企事业单位采购业务相关数据全生命周期安全管控能力方面的技术与管理的脆弱性及威胁性，从而发现自身数据安全问题和短板，明确数据安全保护需求，为数据安全治理的建设指明方向[6]。

1.5 体系建设

为遵循同步规划、同步建设、同步运行的思想，有序落地数据安全防护措施，需在理解合规要求以及紧贴业务场景的前提下，建立数据安全治理体系框架，为政企事业单位构筑以数据为中心的可持续安全治理解决方案。

数据安全治理体系框架包含五个组成部分：以数据安全制度规范体系、数据安全运行管理体系、数据安全技术防护体系形成三维互锁格局。制度规范体系包含顶层设计、流程制度和落地规范等，运营体系包含组织建设、团队管理以及人员能力，技

术体系是实现安全能力的方案、平台、工具等技术能力。再借助数据安全应急响应体系、数据安全监督审计体系进行有力支撑（见图 1）。

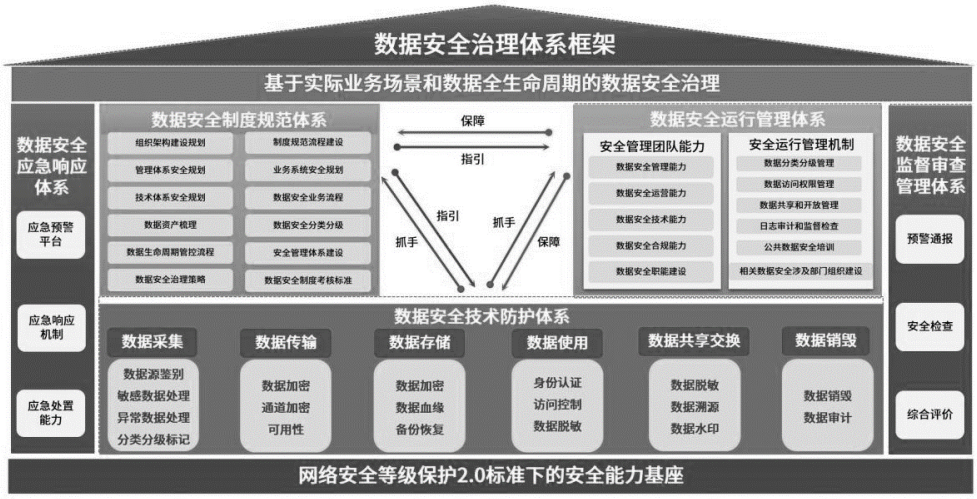


图 1：以数据为中心的安全治理体系

1.5.1 制度体系完善

综合法律法规、业务战略需求及风险现状制定数据安全治理的相关制度及规范，如组织架构建设规划、数据安全业务流程、数据安全制度考核标准等，从而指引技术工具的部署和运营管理建设。

1.5.2 技术工具实施

具体到技术层面，将人员、时间、操作、习惯、应用、网络、介质、主机等多个环节精细管理，为政企事业单位构建可信的安全环境。以现有安全基础设施、等级保护技术措施为基础，并针对性的将数据加密、数据访问控制、数据防泄漏、数据脱敏、容灾备份等多种数据安全技术措施与现有安全防护手段相结合，构建全方位的采购业务相关数据防护体系[7][8]。还可结合领先的大数据分析技术，建立数据安全治理平台，实现数据安全态势感知，全面提升数据安全防护能力，筑牢数据安全防线。

以数据泄露防护系统为例，这套系统由两部分组成：统一管理平台和终端 DLP（见图 2）。统一管理平台部署在政企事业单位信息中心云平台，实现数据梳理及分类分级，集中下策略规则，同时进行敏感数据安全事件监控、处理和统计分析；终端 DLP 则部署在政企事业单位内各计算机终端，发现、识别、监控终端中的敏感数据，对数据资产分布、敏感数据的违规存储进行展现，同时对敏感数据的违规使用、扩散等敏感行为进行策略响应控制。

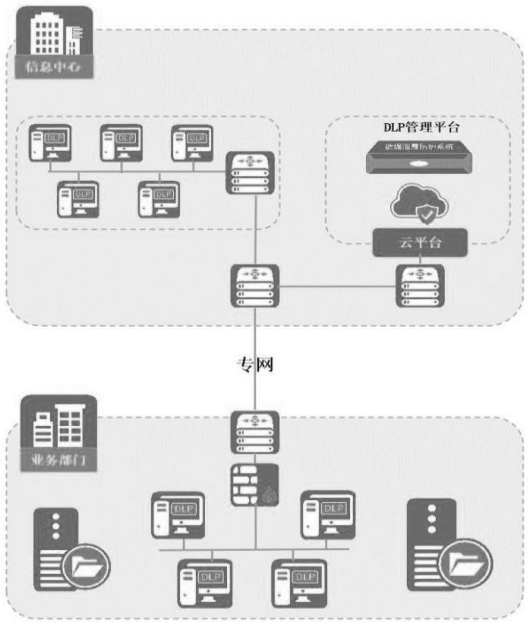


图 2：数据泄露防护系统部署示意图

系统使用后弥补了采购过程信息防泄密工作

在技术层的不足之处；让采购过程信息保密工作实现全方位的管控，从而大大降低泄密风险。

数据泄露防护系统全面梳理和盘点政企事业单位组织内数据资产，对数据全生命周期实施安全保障，集中化数据安全管控策略管理，有效监测数据使用、流转及共享过程中的安全态势及风险，提供了面向数据全生命周期及业务场景的数据安全治理解决方案。

具体来看，在敏感数据的管理规范上，可借助数据泄露防护系统平台，进行数据的分类分级管理，如划分为个人信息类数据、采购类数据、招投标类数据等，基于数据类别进行安全分级，如机密、秘密、绝密，对整体相关敏感文件进行了梳理；同时与政企事业单位的 4A 系统进行对接，基于组织架构和日常工作开展以及办公相关系统的使用情况，基于用户、用户组进行敏感数据的权限管控，定制化数据泄露防护策略，有效防范云盘上传、内部邮箱外发、U 盘拷贝、文件打印及其他社交软件的外发、屏幕截图、拍照等敏感数据泄露风险。

1.5.3 运营管理建设

三分靠技术，七分靠管理。一个好的运营管理体系可帮助企业实现数据安全治理的最终落地与可持续化运转。建立数据安全治理专业团队，提升数安管理、技术、合规能力，加强员工数据安全意识，实现数据安全运营的可视、可控、可持续。

1.5.4 监审与应急体系建设

随着采购业务领域数据安全治理的三个核心体系建设完成，政企事业单位需要针对数据全生命周期的各阶段的安全管理情况进行监控与审计，以保证数据安全治理可以有效、持续地产生价值。在监督审查体系中可以着重于以下几个方向，预警通报、安全监测和综合评价。

针对数据安全事件落实重大事件报告制度和突发数据安全事件应急响应制度，建立健全安全应急预案、应急处置工作指南和处置流程图。常态化开展数据安全攻防演练、应急演练、全员安全培训，组建专家队伍和支撑力量，提升全天候、全场景、常态化、实战化的网络安全应急处置水平。

1.6 持续面向全员的培训

数据安全治理是一项持续的、需要全员参与、全民维护的工程，所以需要提升全民关于数据安全的意识，定期开展相关培训，增强包括数据安全委员会相关人员在内的数据安全知识和能力。

2 方案效果及客户价值

该数据安全治理解决方案可为政企事业单位采购业务领域保密工作提供强有力的支撑，促进部门整体防护能力和规范管理水平不断提升，弥补了采购过程信息防泄密工作的不足之处，让采购过程信息保密工作实现全方位的管控，提高了采购业务人员的保密意识及责任落实，能够对采购活动参与人员形成有力震慑效应，强化采购保密管理的刚性执行，有效杜绝了采购业务信息失泄密事件的发生，持续推动构建“不敢腐、不能腐、不想腐”的长效机制，具备在全国采购业务范围复制推广的重要价值。

参考文献：

[1] 阙天舒，王子玥. 数字经济时代的全球数据安全治理与中国策略[J]. 国际安全研究，2022, 40(1):26.

[2] 龚诗然，刘雪花. 数据安全治理现状研究与分析[J]. 信息通信技术与政策，2022(2):5.

[3] 王林. 新时代我国的数据安全风险及治理方案探析[J]. 山东科技大学学报:社会科学版，2022, 24(3):7.

[4] 贾璐. 数字经济时代数据安全的治理路径探究[J]. 保密科学技术，2022(2):4.

[5] 王庆德，吕欣，王慧钧,等. 数据安全治理的行业实践研究[J]. 信息安全研究，2022, 8(4):7.

[6] 李雪妮，秦书锴. 数据安全治理能力评估框架构建研究[J]. 信息通信技术与政策，2022(2):5.

[7] 张心怡. 数据安全治理体系的构建与实践探索[J]. 大数据时代，2022(6):16.

[8] 高磊，赵章界，宋劲松,等. 大数据应用中的数据安全治理技术与实践[J]. 信息安全研究，2022, 8(4):7.

城市级数据中台数据安全体系的构建

柯杜芹
（泉州大数据运营服务有限公司，福建 泉州 362000）

摘　要：为应对城市级数据中台建设运营过程中存在的数据安全挑战，本文提出数据安全体系的整体构建方案，遵循数据安全能力成熟度模型（DSMM），在管理机制上明确组织建设、制度建设和流程规范，在技术防护上明确数据全生命周期的安全防护技术，以服务数据中台的运维运营，在实践中取得良好的成效。

关键词：数据安全；数据中台；DSMM；数据资产；监测预警

0 引言

随着“大数据”时代的来临和数字政府、数字经济的不断进步，数据作为一种重要战略资源，其作用也日益凸显。鉴于政府掌握着大量的数据资源，这些数据资源由不同的单位和部门负责管理，如果不进行数据的汇聚整合及共享应用，容易造成信息孤岛的现象，难以充分发挥数据的价值。有鉴于此，2015 年以来，国务院相继颁布了《政务信息资源共享管理暂行办法》《加快推进“互联网+政务服务”工作的指导意见》等文件，明确要求建设形成跨部门数据资源共享共用格局。为进一步发挥数据价值，2020 年 5 月，国务院办公厅印发《公共数据资源开发利用试点方案》，确定以数据安全为底线，以“可用不可见”为基本原则，在强化公共数据开发利用安全保障前提下，探索公共数据资源开发利用的机制。

泉州市积极贯彻落实国家、福建省有关政策法规及文件精神，于 2020 年 6 月建成泉州市政务数据汇聚与共享应用平台，通过与省级平台和各自建系统的对接，实现国家、省级泉州属地政务数据的回流和市、县两级自建系统业务数据的汇聚，建立全市政务数据资源中心；通过统一的数据共享服务方式，打破“数据壁垒”，实现跨层级、跨部门、跨系统的数据共享需求。2021 年底，泉州市进一步建设上线泉州市公共数据资源开发服务平台，为我市公共数据资源开发利用提供基本的运行环境。市汇聚共享平台和开发平台为全市提供统一的数据汇聚存储、清洗治理、共享开放、开发利用等能力，初步构成城市级数据中台。

随着《数据安全法》《个人信息保护法》的颁布实施，数据安全问题成为数据应用过程中不可忽视的主要环节。随着平台的数据总量越来越庞大，有必要针对城市级数据中台构建统一的数据安全体系，为数据的全生命周期提供安全保障。

1 数据安全面临的挑战

与传统的单个信息系统存储的数据资产总量较为有限、数据安全责任较为清晰相比，城市级数据中台由于对接多个单位多套信息系统的数据汇聚与共享应用，平台的数据总量越来越庞大，其中不乏涉及个人隐私数据和政府重要数据，这类数据均有极高的数据加工及买卖价值，由于数据集中存储，且人群高度覆盖，成为了极具价值的数据资产。这些数据资产一旦被篡改或者泄露后，造成的社会风险难以估量。如何做好这些数据的安全保障，从各个层面防止数据被攻击、遭到泄露成为了城市级数据中台亟待解决的问题。经梳理，平台数据安全面临的挑战主要有如下几点：

（1）安全管理方面的挑战

平台数据资产总量庞大，涉及多个数据流转环节，每个环节的数据处理及运维人员各不相同，系

统使用、运维等存在的数据安全隐患较大，如未建立有效的数据安全管理机制，容易出现监管处置盲区、发生数据安全事故无法明确具体责任人等问题。

（2）资产管理方面的挑战

针对平台数十亿或上百亿的数据总量，数据资产的盘点是一项巨大繁杂的工程。如未进行有效的数据资产管理，容易出现数据资产分布情况不清晰、存在被忽视遗漏的数据、不清楚哪些属于应重点保护的敏感数据等问题。

（3）流转使用方面的挑战

平台的数据涉及汇聚存储、清洗治理到共享开放、开发利用等数据全生命周期的各个环节，任何一个环节均存在数据被篡改、失泄密等安全隐患。如未采取有效的安全防护措施，并进行有效的数据流转使用记录，容易出现严重的数据安全事故。

（4）应急处置方面的挑战

平台承载着整个城市的重要数据资源，一旦发生数据访问操作使用方面的异常，如越权访问、超正常使用量请求等，如无法有效进行自动告警及应急处置，容易进一步扩大问题影响面，带来进一步的严重后果。

2 总体设计思路

城市级数据中台数据安全体系整体遵循 GB/T 37988-2019 定义的数据安全能力成熟度模型（DSMM），即由以下三个维度构成：

安全能力维度明确了组织在数据安全领域应具备的能力，包括组织建设、制度流程、技术工具和人员能力。

能力成熟度等级共分为 5 级，具体包括非正式执行级、计划跟踪级、充分定义级、量化控制级和持续优化级。

数据安全过程具体包括数据生存周期安全过程（数据采集安全、数据传输安全、数据存储安全、数据处理安全、数据交换安全、数据销毁安全）和通用安全过程。

DSMM 架构如下图所示：

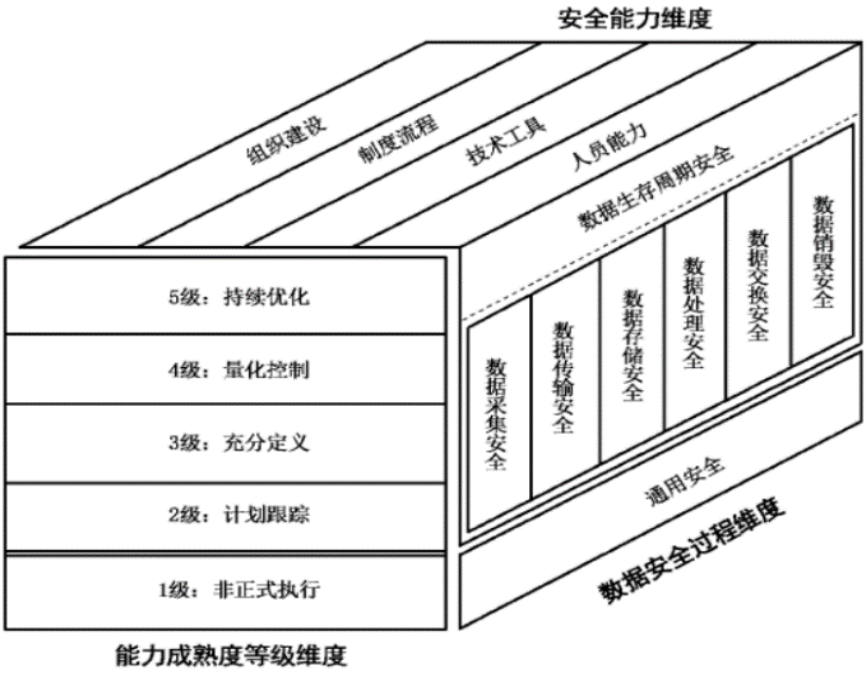


图 1　DSMM 架构图

数据安全体系总体架构设计如下：

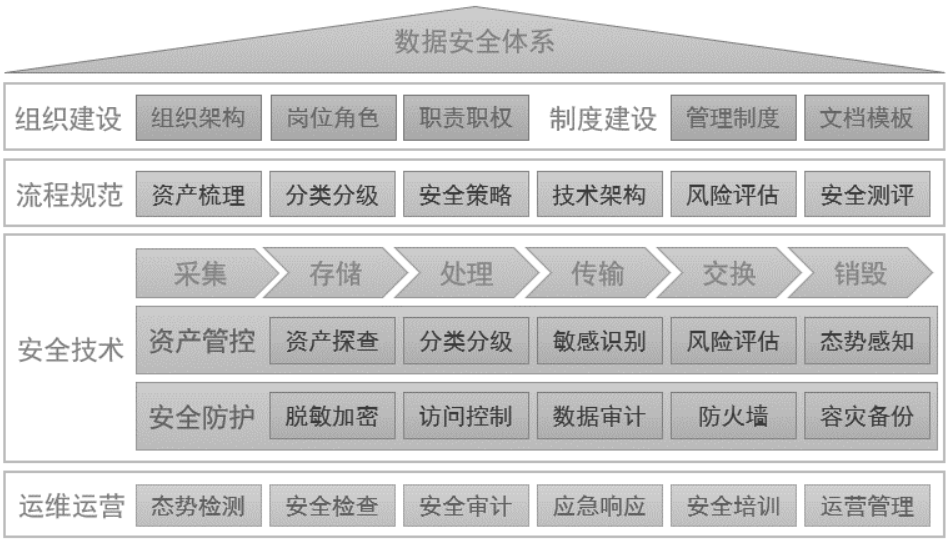


图 2　数据安全体系总体架构图

概括地讲，就是在以下几个方面持续进行优化：一是从战略规划、组织建设、人员配备、管理制度等方面，不断建立健全数据安全管理体系；二是持续开展数据资产盘点和分类分级，包括资产梳理、标签处理等；三是定期开展数据安全风险评估，针对安全合规情况进行对标分析，对风险漏洞进行筛查，对安全隐患进行整改；四是做好数据全生命周期的安全防护，针对不同级别的数据配置部署相应的防护策略和管控措施；五是做好数据安全运维运营管控，通过统一管理、统一策略，做好安全检查和监督整改。

在数据生命周期安全方面，关键的安全防护措施如下图所示：

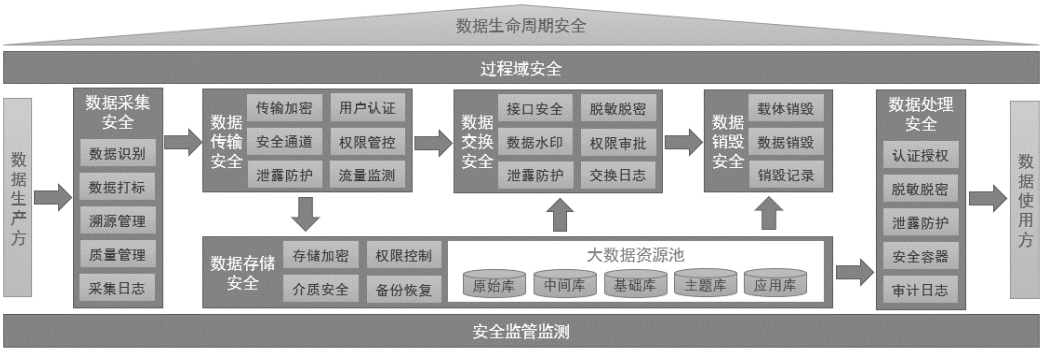


图 3　数据生命周期安全防护图

3 具体实施方案

（一）管理机制

一是健全管理组织架构。按照“政企协同、管运分离”的机制，由大数据主管部门负责城市级数据中台数据安全体系的业务指导和监督检查，成立国有全资大数据公司，负责城市级数据中台的建设运营及数据安全防护的具体落地实施。大数据公司内部成立数据安全领导小组，明确数据中台的数据安全具体责任部门和责任人，落实具体安全防护策

略的执行。通过大数据公司进行专业化运营、大数据主管部门进行定向采购服务的机制，有效解决大数据主管部门人手不足、难以聚焦具体技术实现及运营实施等难题。

二是完善数据安全规章制度。大数据公司在大数据主管部门的指导下，起草数据中台数据安全相关的各项规章制度，并按规定实施，具体包括：《数据中台安全保障体系规划》《数据中台安全管理办法》《数据分类分级指南》《数据共享安全管理规范》《数据中台运行监管日志规范》《数据中台数据授权管理规范》等。

（二）技术防护

在城市级数据中台建设运营的基础上，建设统一的数据安全监管监测子平台，实现数据中台的数据安全全面保障，辅助管理人员进行数据资产的管控，防止数据遭到泄露。数据安全监管监测子平台的主要功能和成效如下：

（1）数据资产分析管理

通过数据安全监管监测子平台对数据中台上的各个应用子系统进行定期主动的数据资产扫描，实现对数据中台业务数据库资产自动发现及数据分级分类管理，帮助数据中台管理员掌握数据资产的分布、使用状况以及数据库分布。相关资产梳理结果可按周期统计结果生成报告，并可对比分析不同周期内的差异性变化，从而让管理员全方位的了解资产的分布。

基于建立的数据资产底账，辅以人工服务中刻画的管理域边界，通过对数据访问行为的分析，动态侦测数据资产的变化情况，发现数据中台中存在的僵尸资产、复用资产、未知资产、高频资产、失踪资产、销毁资产和不明资产，从而帮助管理员快速刻画指定责任范围内具有风险的数据资产。对于明确的僵尸资产，支持由管理员进行确认并一键删除。而对于无用的僵尸数据库账号，同样支持由管理员进行确认并一键删除，以避免因为疏于管理而导致被非法利用。对于高频资产进行刻画管理，帮助管理员熟悉明确哪些数据是频繁使用的，需要重点防护。

数据中台数据资产发现分析情况（样例数据）如下：

图 4　数据资产发现分析示例图

（2）用户权限监测分析

通过数据安全监管监测子平台，对数据中台上的数据库账号及权限进行如下监测分析：

一是对数据库账号进行监测，对数据库账号的终端登录情况全面了解，避免数据库账号发生泄漏却不知，导致非法用户登录恶意修改，同时避免数据库账号复用导致安全责任人难以定责等安全风险。

二是对数据库账户口令进行监测，避免出现弱口令导致数据库极易被攻破，从而造成存在巨大安全隐患。

三是对数据库用户权限进行分析，提出将用户

权限在其职责范围内进行最小化处置的建议，从而避免用户权限过大导致安全事件的发生。

（3）数据安全监测预警

通过数据安全监管监测子平台，利用探针监测、人工智能算法等技术，对数据中台上的数据库服务进行如下安全监测预警：

一是当数据库主机存在主动访问外部流量的现象时，在排除掉数据库定期备份等正常现象之后，自动确认是否有可能是数据库已经遭受到了外部控制，并及时自动告警及处理。

二是自动发现是否有开放非数据库的服务，可找到非数据库的服务并及时关闭，以避免提供外部攻击的通道。

三是当数据中台数据库中存在持续时长较长的语句时，进行进一步分析以确认是否有可能是外部攻击行为导致，并及时自动处理。

四是对数据中台相关应用系统访问数据库的行为进行分析，根据最多的访问行为分析出正常访问行为特征，在后续的访问行为中如出现被判断为异常访问时，将进行核实处理。

五是对数据库服务流量进行分析，当服务流量在短期内出现暴增时，有可能是遭受到外部攻击，将自动进行进一步判断并采取相应的安全措施。

（4）行为审计与追根溯源

通过数据安全监管监测子平台，对数据中台上的数据库操作行为进行监测。平台完整记录对数据库的所有操作，以便用户在未知的风险事件发生后，定位问题的发生过程。平台可实现在以亿为单位的数据中，多条件查询数据，在数秒内返回结果，同时对海量数据实现压缩比 90%以上的高性能存储。平台提供多维度海量审计数据对比分析工具，从不同的空间、时间对各个维度进行同比和环比分析。

平台通过完整记录用户访问的所有操作，在未知的风险事件发生后，可快速定位问题的发生过程进行追根溯源。以安全事件为入口，以风险模型为基准，针对事件全过程中的任意线索进行多维拓展，利用云端丰富的实时威胁情报和本地的网络行为、终端行为以及多维度海量审计数据对比分析，从不同的空间、时间对各个维度进行同比和环比分析，对安全事件进行回溯和调查，可视化绘制出完整的事件生命周期，包括的源、目标、途径、范围等相关信息，为平台的数据库的安全防范提供支撑。具体追踪溯源（样例数据）如下图所示：

图 5　事件追踪溯源分析示例图

（5）数据应用安全防护

通过数据安全监管监测子平台，对数据中台所有对外服务的数据应用接口安装流量探针，抓取所有的访问行为记录，进行安全管控。通过大数据智能分析算法识别访问目标是页面调用，还是接口访问，将结果与已知的接口列表进行匹配，发现是否

存在未知接口。同时对接口数据进行深入分析，识别具体传输的数据字段，以确保是在授权范围内的数据获取和通信。

针对数据开发利用需遵循数据“可用不可见”的数据安全保密需求，平台提供数据安全计算沙箱环境，通过对计算沙箱配备必要的安全隔离防护技术，实现不同数据沙箱实例之间相互不可见，并对离开沙箱环境的运算结果数据进行必要的监控，确保原始数据不出数据中台、结果数据不可逆。

（6）运维态势安全研判

通过数据安全监管监测子平台，实现对运维态势的监控，通过运维态势大屏，实时掌握运维状态，分析运维人员的账号，是否权限过大，是否存在僵尸或复用账号等，防止从内部运维引发安全事件的风险。提供运维态势视图全面展示用户整体运维态势情况，并可指定周期进行具体运维情况的查询。具体态势感知情况（样例数据）如下图所示：

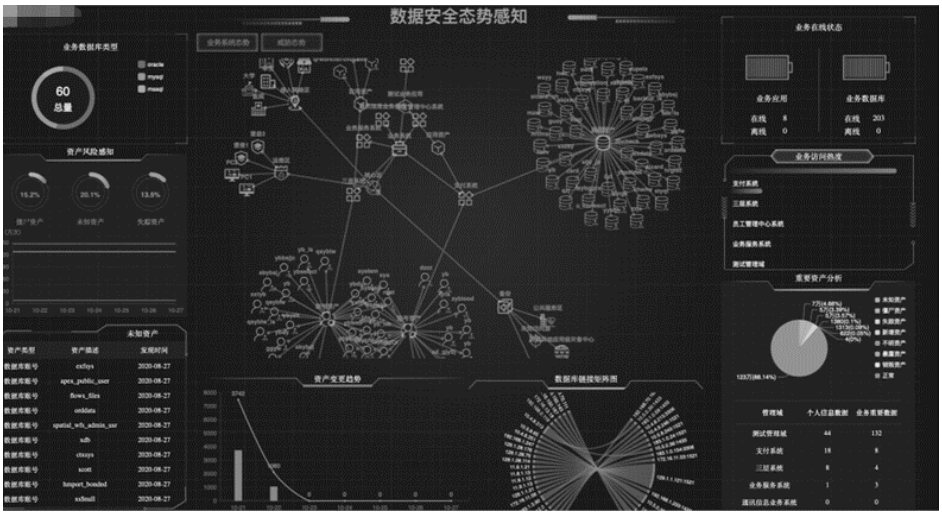


图 6　数据安全态势感知大屏示例图

5 总 结

城市级数据中台数据安全体系伴随着数据中台的上线同步开展实施并不断优化完善，截至 2022 年 7 月，平台数据总量已突破 60 亿条，目前日均新增汇聚数据量在 300 万条以上，发布数据共享服务接口 350 个，提供 2000 多项数据批量交换服务，交换总记录数达到 18.8 亿条数据，为疫情防控、普惠金融、应急管理等相关应用场景提供数据应用支撑，整体运行状态安全平稳，数据资产分类分级梳理及全生命周期监管有效落实到位，尚未发生重大数据安全事件。

未来的数据安全体系将在充分注重基础防护的基础上，聚焦于更为智能、全面的自动化分析处置能力提升，包括引入人工智能、大数据、区块链等算法技术，以进一步减轻安全运维管理人员的日常工作负担。

以微隔离之名，行“零信任”之事

俞志荣
（麦讯天下（福建）信息技术有限公司，福建 福州，350001）

摘　要：传统的网络安全模型基于边界防护的思想，无法适应当前的需求.零信任是一种新的网络安全模型，能够用于防护边界日益模糊的网络. 而微隔离是零信任的结构性要求,本方案介绍了微隔离的定义，分析其创新点，并总结了发展现状，可为微隔离的研究与应用提供参考。

关键词：零信任；微隔离；网络安全

1 微隔离诞生的背景

1.1 数据中心云化带来的安全挑战

传统的安全产品基本都是在南北向业务模型的基础上进行研发设计的，从物理机到虚拟化，再到容器技术,尽管十几年间数据中心技术发生了翻天覆地的变化,但遗憾的是安全防护的思维却未有实质进展——注意力始终在外部边界的安全,而很少关注内部究竟发生了什么。

1.1.1 数据泄露持续发生

尽管数据越来越受到重视,相关数据保护的法规不断增多，企业、政府组建了专业的安全团队，也增加了数据保护方面的安全投入,但数据泄露事件仍持续发生，而这些攻击都有一些显著特点，一旦边界的防线被攻破或绕过,攻击者就可以在数据中心内部横向移动,而中心内部基本没有安全控制的手段可以阻止攻击。

1.1.2 病毒泛滥

从以破坏为主，到隐藏控制，再到传播勒索，攻击者利用病毒的攻击逐渐具备了更加明确的政治或经济目的。据调查，企业服务器被勒索病毒加密的事件在 2018 年上半年增长了 34%，受害企业更偏向传统行业,前三位分别为政府机关（26%）、工业企业（15%）和医疗机构（13%）。

1.1.3 横向侧移攻击

目前的网络攻击主要利用了网络安全设计中的弱点：一旦突破边界的防护，内部基本没有访问控制的手段。从著名的洛克希德-马丁提出的网络攻击杀伤链来看在数据中心攻击链中,攻击者依赖内部的横向转移来扩大攻击范围,从而达到攻击目的。如果内部没有有效的隔离及安全控制，面对云数据中心中大量虚拟机的复杂内部访问,安全人员几乎无法发现内部存在的攻击。

1.2 微隔离的需求痛点

东流量不可视:服务器之间的流量往往不可见，至少不能完整的洞察，在微隔离诞生之前，几乎没有这样的技术和方案,这就导致企业无法准确的制定安全策略，正所谓“看不见、没法管”；

静态策略不可用:当前的大部分云数据中心内部，工作负载的“漂移”属性特别强，尤其是容器化之后，一个容器的消亡、产生、扩缩容，都会产生新的 IP 地址，传统的静态策略显然是跟不上工作负载动态变化；

人工运维不可行：由于工作负载规模庞大，东西向流量巨大,要精细化控制是一件很复杂的事情,依靠手工运维显然不太可能,就算可能这个过程对于运维人员来说也是十分的痛苦。

混合架构无法管:国内数据中心大多数存在异构、多云、混合云等特点，造成安全策略迁移、维护与管理上的困难。

2 解决方案

2.1 微隔离的定义

2016 年 Gartner 副总裁，知名分析师 Neil Mac-Donald 在 Gartner 安全域风险管理峰会上提出微隔离概念。微隔离技术被 Gartner 连续 3 年评为全球十大安全项目之一，并在 2018 年的《Hyper Cycle for Threat-Facing Technologies》中首次超过下一代防火墙。

- Gartner 对于微隔离的定义

微隔离（也称为软件定义隔离）使用策略驱动的防火墙（通常基于软件）或网络加密来隔离数据中心，公共云和容器中的工作负载，包括混合和多云场景中的工作负载和跨越所有这些场景的工作负载。

- CCRC 对于微隔离的定义

2018 年，中国网络安全审查技术与认证中心发布国内第一个《微隔离产品安全技术要求》，该标准将微隔离定义为“一种能够适应虚拟化部署环境，能够识别和管理与云平台内部流量的一种隔离技术”。

2.2 微隔离与零信任的关系

零信任作为一种指导企业构建安全网络环境的模型和思想，完全实现零信任需要确保用户终端可信、用户可信、链路可信、网关可信、应用资源可信。零信任的核心原则：“永远验证，绝不轻信”，基于这三大原则，常见的零信任方案分为三种：

基于身份治理的零信任

基于微隔离的零信任

基于软件定义边界的零信任

从国内外的落地的零信任项目的实践来看，企业可以根据自身的组织架构、业务流程及工作场景引入不同产品组件构建自己的零信任方案，往往企业在选择的过程中会发现，不同的实现方案意味着不同深度的效果。因此狭义的角度来看，微隔离是构建整个零信任安全框架最为核心、必不可少的组成部分，是零信任的结构性要求。

2.3 微隔离的关键功能

2.3.1 可视化业务拓扑

东西向流量的可视化管理是现代数据中心管理最为迫切的需求之一。由于当代数据中心普遍具有计算节点众多，内部流量复杂的特点，使得东西向流量很难被准确的理解，进而也就无法被有效的管理甚至优化，因此，迫切需要提供一种技术能力对东西向流量进行清晰的展现和梳理。

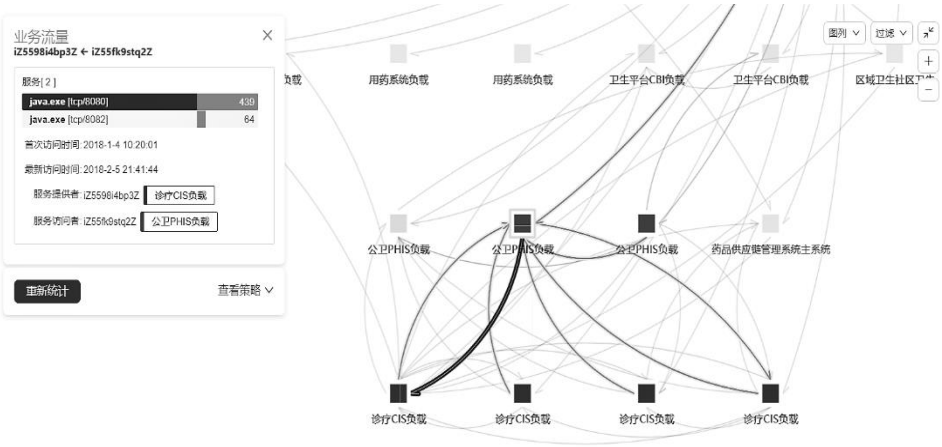


图 2.3.1 可视化示意图

2.3.2 面向业务的策略管理

现代云数据中心安全运维复杂，策略不易调整的很大一个原因就是沿用了传统安全运维中基于 IP 的管理体系。反观现代数据中心，动辄几百上千的虚拟机，内部安全管理让管理者望而却步。

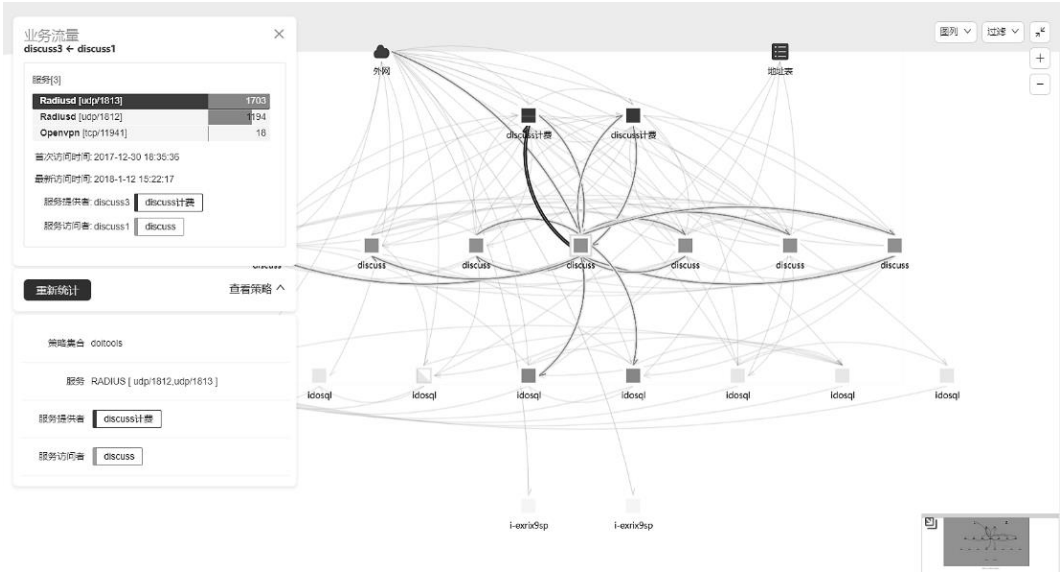


图 2.3.2 面向业务的管理策略示意图

2.3.4 自适应微隔离

数据中心的一个重要的管理场景就是应用迁移及拓展，可能有很多原因触发这一过程，比如从物理机迁移到虚拟机，新的数据中心建设，新的业务上线，应对短期流量高峰等。使用传统基于网络的解决方案，往往限制了敏捷性，而且并没有满足保护应用程序和数据的安全要求。

2.3.5 基础架构无关

数据中心由多种底层技术架构混合而成的，而实现形式也多种多样，有私有云、公有云、混合云等。用户往往很难找到一种能够跨越所有这些技术架构的管理手段来对数据中心进行统一的安全管理方式。

3 微隔离在国内外的发展状况

3.1 国外发展情况概述

微隔离一词最早由 VMware 的 NSX 团队提出并定义为：一种在数据中心和云部署中创建安全域起到隔离工作负载并进行个别防护效果的技术，其目标是实现更细粒度的网络防护。

NSX 网络虚拟化将服务器虚拟化技术移植到了虚拟化网络架构中，NSX 使用类似于 Hypervisor 的“网络虚拟化管理程序” 再软件中重现二到七层的整套网络服务。数据中心搭建 NSX 虚拟化网络平台后，对于管理员来说，所有虚拟机就如同连接在一台物理交换机上一样。

另外，illumio 公司的自适应安全平台是国外较为成熟的微隔离技术，该软件运行在较高特权级（可与内核模块相结合），可以捕获系统运行时的状态，截获所有出入的流量，阻止恶意程序的执行等等。

3.2 国内发展情况概述

国外尤其以美国为代表的微隔离发展较早，用户接受度高，而国内微隔离技术则还处于发展阶段。其中流量牵引方案是国内技术比较成熟的微隔离方案。基于引流的微隔离方案的实现方式是通过将所有虚拟机的网关设定为同一台虚拟机（或物理主机）将其作为网关，这样所有的虚拟机流量都会经过网关过滤，入站出站的所有流量只有符合过滤规则才能通过。

国内另一种应用比较广泛的微隔离技术通过配置交换机的网络流量过滤功能来实现。将虚拟网络划分为多个隔离域，对域以及域内主机进行细粒度隔离。此种方法优点在于虚拟交换机不仅可以满足负载均衡与冗余，同时也能够对网络设备数量进行缩减，简化网络架构，以此缓解系统管理维护。此种方案实施依然存在一些难点，首先，对虚拟机的防护策略配置问题，大量的虚拟主机如何才能快速配置对应的防护隔离策略。再者，隔离组的划分

并没有一个统一标准，采用何种策略能够尽可能高效。最后，如何集中管理。

目前国内的微隔离技术的发展趋势：

一是实现数据中心不同介质主机（物理机、虚拟机及容器）统一管理的研究，满足内部不同云环境、不同虚拟化架构、不同介质统一管理的需求；

二是实行策略自适应技术研究，有效利用基于内部业务拓扑所形成的访问控制策略的优势，实现数据中心内部主机动态迁移时策略自适应配置更新。

3.3 微隔离技术的创新性

3.3.1 数据中心内部东西向流量可视化

基于全量式和增量式相结合的数据中心内部东西向流连接关系及主机数据的采集方法，利用客户端软件抓取数据中心内部物理主机、虚拟主机及容器间的主机信息及主机间的业务访问关系，快速准确的数据分类和萃取方法，构建业务逻辑拓扑，形成数据中心内部业务关系可视化模型。实现数据中心内部高效率、可交互的显示和获取工作负载信息和工作负载间业务逻辑关系，降低数据中心内部网络管理者的难度及减轻数据中心内部网络管理的工作量。

3.3.2 基于角色属性的策略模型

主机的业务角色决定了其能够向其他主机提供服务的范围，也决定了它所能够访问主机的范围，因此基于角色角色的访问控制模型，“自顶向下”及“自底向上”的访问控制策略构建方法，通过为多维标签形成的角色属性指派具体出站、入站访问连接权限，实现角色属性与权限属性的映射，从而大大简化访问控制策略模型复杂度，降低数据中心内部大规模工作负载节点情况下访问控制策略的设置难度。

3.3.3 防护策略生成及自适应更新技术

基于角色属性的策略管理模型对所有主机划分服务应用角色，利用访问控制策略管理框架预定义具体的安全访问控制策略，及自适应动态调整技术实现服务对象、地址对象等多个维度组合为一体的点对点白名单式东西向访问控制策略，实现数据中心内部域间隔离、环境隔离、应用隔离、主机隔离、甚至端口级隔离。

同时对环境参数持续监控，由统一的计算中心动态调整安全策略，实现策略自适应。

4 微隔离的价值

4.1 核心价值

就目前的实践来看的确有很多企业并没有意识到隔离的重要性，在此之前他们完全没有办法洞察内部的流量，而梳理后的资产和连接关系，则能够直接帮助他们提升管理的能力和水平。很多客户在云化数据中心建设初期就部署了主机加固类产品，但在东西向流量的访控手段上是缺失的，基于微隔离系统提供的分段能力，无论是从数据中心落地零信任的角度，还是从数据中心基础架构加固的角度而言，都对客户收缩暴露面、防护攻击侧移、勒索病毒传播提供了关键能力。

如果黑客已经攻进了一个服务器，那么他就可以利用这个服务器做跳板，进一步攻击网络中的其他服务器。这正好符合了零信任的原则：

（1）假设已经被攻破

（2）持续验证，绝不轻信

（3）只授予必须的最小权限

而微隔离的核心价值在于：微隔离可以阻止这种来自内部的横向攻击。微隔离通过服务器间的访问控制，阻断勒索病毒在内部网络中的蔓延，降低黑客的攻击面。

4.2 普适性价值

- 协助满足部分等保 2.0 规定

根据国内特有情况，微隔离产品可以满足等保2.0《安全通用要求》与《云计算拓展要求》的共 21 项子要求；

- 减少数据中心内部被攻击风险

主机每一个开放的端口就是一个攻击入口，而每一个能够访问这台主机的计算节点都是一个可能攻击源。而安全管理的一个重要方式就是减少攻击面，基本做法就是一方面关闭不必要开放的服务端口，一方面限制能够访问主机的来源。

- 安全策略集中可视化管理

摆脱传统命令行的安全运维方式，可通过业务

拓扑图，在查看业务实际访问逻辑的基础上进行策略配置，并使用基于 ID 代替基于 IP 的策略管理方式，可大幅度缩减安全策略总数。

- 混合云统一安全管理

企业从业务可用性、安全性及可靠性等角度考虑，往往会采用多个云平台构建自己的混合云，而安全往往是割裂的，每一个云平台均有自己的一套安全运维工具及产品，管理人员通常只能采取较为宽松的管理方式，从而留下了较大被攻击的风险。

主流微隔离产品面向操作系统开发，能够提供跨平台（任何基础架构的公有云、私有云、混合云）、跨介质（物理机、虚拟机、容器）的统一安全管理能力。

5 微隔离产品目标群体

如同大多数的新技术一样，微隔离技术短期内大范围的普及是不现实的，毕竟“新”的技术都存在一个缓慢接受的过程。因此微隔离目标群体包括但不限于存在以下特点的用户：

- 数据中心环境复杂：多种云平台、多类操作系统、虚拟机与容器均具备，环境越复杂微隔离需求越强、微隔离的可替代性越弱；
- 业务运行于云原生环境：规模化容器部署情况下，微隔离需求刚性，云原生程度越强、微隔离需求越迫切；
- 已部署 CWPP 产品：企业重视数据中心云工作负载保护，网络控制能力亟待补足；
- 严重的安全事件：如勒索病毒爆发、工作负载大面积失陷、核心业务被攻击等；
- 新技术的探索任务：安全建设接近饱和，对创新技术赛道兴趣浓厚或接受程度高；

综上，微隔离技术目前在国内的目标用户行业主要集中在金融、互联网、国企/央企、大型制造业等具备较高安全要求的行业。

北卡密甲：基于国密算法的工业互联网数据安全解决方案

翁才杰　林幸华　邱丽灵　阮莉丽
（卡科技有限公司，福建 福州 350108）

摘　要：首先介绍基于国密算法的工业互联网数据安全解决方案——北卡密甲面向的客户群体，其次通过分析现有工业互联网数据安全存在的问题，从系统架构、技术难点、功能特色几个方面对北卡密甲进行介绍，并阐述了建设北卡密甲对工业互联网行业的支撑作用，最后介绍了北卡密甲的创新性与先进性。

关键词：国密算法；工业互联网；数据安全；北卡密甲

1 目标客户群体

政企事业单位、互联网、电信、金融、医疗、工业互联网等行业或大型企业。

2 解决方案拟解决的问题

2.1 方案综述

数据是国家重要的基础性战略资源，也是企业/机构的宝贵财富，2019 年 7 月，工信部等十部委联合发布了《加强工业互联网安全工作的指导意见》，明确要“强化工业互联网数据安全保护能力”。

数据既是工业互联网的核心要素，也是安全薄弱环节。工业互联网数据分布在大数据平台、用户端、生产终端等多种设施上，设备种类庞杂、位置分散、缺乏安全设计；高端设备、高端 PLC 器件等依赖国外发达国家，重要工业资产和装备制造信息可能被国外非法收集；基于开放与标准化的原因，工业互联网也越来越多地使用公开协议以及标准化的 Windows 或 Unix 技术架构，其安全漏洞使攻击门槛大为降低。因此，工业互联网数据存在被泄露、窃取、伪造，以及被利用来攻击等严重安全风险，对数据安全威胁的及时感知与防护十分有必要。数据安全已成为工业互联网的主要安全技术瓶颈。

因此，亟需引入以身份为中心的零信任框架，研发基于国密算法的自主可控的数据安全技术，包括身份认证、数据加密、安全传输，基于行为分析实时感知异常的技术，为工业互联网数据安全赋能。

2.2 核心价值

北卡密甲是基于国密算法的工业互联网数据安全解决方案，综合采用基于零信任安全的身份认证与鉴别技术、基于国密算法的数据安全传输技术以及工业互联网的安全行为智能分析方法，在接入、传输、管理等数据安全的关键环节构建具有内生性的安全防护能力。北卡密甲通过安全网关对接工业互联网，工业互联网中的设备和用户将利用安全网关进行身份认证和加解密，实现工业互联网数据安全传输。

2.2.1 系统架构

北卡密甲通过外置方式（或内嵌 SDK 方式）安全网关以最小代价接入工业互联网，不改变原有网络总体架构。

图 1 北卡密甲结构拓扑图

2.2.2 技术难点

（1）算法效率要求高

与消费网络不同，工业互联网中连接的是工控系统，工控系统往往需要进行机床控制指令和关键生产数据的实时传输，对实时性的要求极高，在工控系统中简单集成普通的密码算法会影响系统的功能性，消耗设备的计算能力。

（2）对外置设备的硬件要求高

工业互联网的底层是物联网，工业互联网中数据的采集与传输主要依赖物联网，物联网对设备的体积和功耗有比较高的要求。同时，工业互联网中的硬件设备的运行环境是工业生成环境，这就要求硬件设备要能够在高温、高压、电磁干扰等复杂环境中安全稳定长期运行。

（3）方案兼容性要求高

接入工业互联网的设备形式多样，功能差异很大，有些简单终端设备通常硬件配置不高，仅能支持有限的网络接入方式。而工控系统应用程序和协议最初在设计开发时并未采用认证和加密机制以及其他安全策略来防护系统安全，在工业互联网中有些简单功能的终端也需要承担高安全性要求的业务。但因其自身资源和计算能力有限，无法使用常规的方式进行身份认证，并且进行身份认证还需要兼顾设备接入的效率要求，同时由于工控系统本身具有关键性和敏感性，还需兼顾改造与升级可能导致的未知危险。

2.2.3 功能特色

（1）私有化部署

用户可将北卡密甲安全网关部署在自己的环境中，进行嵌入开发与独立运维。

（2）高效集成

用户可直接调用安全网关，无需更改原有系统业务逻辑，可快速为系统增设安全铠甲。

（3）可信认证

基于国密算法的 PKI 和 IBC 组合身份认证方法，确保系统与终端身份合法，防止中间人攻击。

（4）两层加密

在通信双方的安全网关之间建立传输层安全通道，在应用层端到端加密，实现数据安全传输。

（5）国密算法

采用私有通信协议与国密算法（SM9、SM3、SM4、ZUC 等）。

（6）会话密钥

采用国密算法的密钥协商机制，会话密钥仅由通信双方掌控，“一话一密”防止重放攻击。

（7）多媒体加密

支持文字、文件、图片、音视频流等信息的高效加密，确保数据流安全。

2.3 建设价值

2.3.1 数据安全是工业互联网的核心技术

工业互联网的架构主要由三大支柱组成：数据、网络、安全。从防护对象的视角看，工业互联网安全架构涉及设备、控制、网络、应用、数据等五类对象，在具体技术层面，都将落实到一个共同的对象，即数据。保护数据，就是在保护工业互联网产业。

2.3.2 数据驱动改变工业互联网安全范式

目前国际上公认的，解决网络安全攻防不对称

问题的方法之一，就是数据驱动安全。数据既是保护的对象，又是可以用来提供安全防护的工具。在数据驱动思想下，采集设备、网络数据，运用人工智能、大数据分析、基于协议深度解析，以及事件关联分析等技术，可分析工业互联网当前运行状态并研判安全态势，并助力工业互联网安全。

2.3.3 数据安全是工业互联网内生性的安全防御能力

发展内生性的安全防御能力是未来工业互联网安全防护的技术趋势。对数据进行加密和访问控制，是工业互联网重要的内生性安全防御能力之一。数据的访问、传输具有动态性、实时性，是极易出现安全问题的环节。借助数据加密、身份验证、访问控制、完整性验证等机制将有效提升数据流转的安全性，将使得工业互联网数据得到安全高效的利用，并提升工业互联网整体安全性。

本方案通过工业互联网数据安全关键技术保障数据安全访问、传输与使用，响应了工业互联网安全产业的核心诉求，对工业互联网设备安全、网络安全、平台安全等方面也都有着重要积极作用，将与其他工业互联网安全技术，协同支撑工业互联网安全生态体系建设。

3 方案的创新性与先进性

3.1 方案的创新性

3.1.1 零信任架构下的工业互联网数据安全保护体系

近年来，网络安全模型正在从基于节点与边界安全的网络中心化模型，转变为基于认证和授权的身份中心化的零信任模型。零信任安全摒弃传统网络安全的防护架构，整合现有成熟技术，不断兼容新技术，根据行业用户业务需求及安全需求、用户所处地理位置等，对用户进行分类分级管理，并根据相关标准规范，对每个组的用户策略进行设置，是一种有效维护行业内部数据安全的新型网络架构，但在工业互联网领域的应用尚属于起步阶段。本方案在零信任安全框架下，通过国密算法在工业互联网数据传输中的应用，结合可信身份认证与安全接入、安全行为智能分析，从而实现工业互联网的数据安全保护方案，可望探索工业互联网数据安全新的研究思路。

3.1.2 基于国密算法的工业互联网数据安全保护方案

国密算法是由我国密码管理局颁布，自主研发创新的一套商用数据加密处理系列算法，处于完全指数级计算复杂度，相同安全等级下使用公钥位数少，具有更强的安全性和更高的效率。国家有关机关和监管机构站在国家安全和长远战略的高度，陆续提出推动国密算法应用实施、加强行业安全可控的要求。2019 年 7 月，工信部、教育部等十部委联合发布的《加强工业互联网安全工作的指导意见》中也明确提出，鼓励商用密码在工业互联网数据安全保护工作中的应用。目前虽然有机构起草了工业互联网信息安全相关的标准，但是国密算法在工业互联网中的应用还比较鲜见。本方案结合在国密算法领域多年的研究经验，研究 PKI 和 IBC 的组合实现工控设备和终端设备鉴别；采用基于椭圆曲线的国密安全密钥交换协议进行密钥协商，通过端到端加密技术确保数据传输安全，并进一步结合国密 SM2 非对称加解密算法，结合信息隐藏技术、信道隐写等方法进行隐蔽安全传输，实现关键信息传输的双重保护。这些技术将进一步拓展国密算法在工业互联网的应用，有效保护工业互联网数据安全。

3.1.3 人工智能技术在工业互联网安全的应用

目前，人工智能技术主流用于网络安全，而对工业互联网安全应用甚少。工业互联网实现了全系统、全产业链和全生命周期的互联互通，而与此同时，互联互通的实现也打破传统工业相对封闭可信的生产环境，导致攻击路径大大增加，数据种类和保护需求多样，数据流动方向和路径复杂。因此，利用人工智能技术发现数据中的通信行为模式，检测异常、预测安全趋势是工业互联网中的先决条件。本方案利用机器学习、模糊推理、证据推理等方法智能分析工业互联网安全行为，构建态势感知系统，有效提高预警准确率，同时动态更新系统可信用户，这些技术将推动人工智能技术在工业互联网安全的“新战力”。

3.2 方案的先进性

3.2.1 构建基于国密算法的 PKI 和 IBC 组合身份认证方法

基于国密算法的 PKI 和 IBC 组合身份认证方法是零信任安全架构下身份认证和安全接入的关键环节。安全身份认证系统的身份认证与鉴别采用的是国密算法中的非对称密码，将基于 PKI 体系的 SM2 数字证书的强签名，实现跨系统、跨域的身份认证和授权管理；使用基于 IBC 的 SM9 密码算法，利用标识密码系统中每个实体具有一个有意义的、唯一标识的性质，将标识本身作为实体的公钥，使用过程无需预先协商密码或者交换证书，减少传统证书体系中申请和验证环节，降低密钥和系统管理成本。该技术为 PKI 应用到工控系统做了良好铺垫，增强工控系统身份鉴别的安全性，从而实现本方案的身份认证和安全接入。

3.2.2 基于国密算法加密体系实现工业互联网数据传输

本方案采用 SM 系列国密算法，使用 SSL/TLS 技术建立安全传输通道。通信双方的每一次通信前，通信双方的通过各自的标识信息采用基于标识的国密 SM9 算法的密钥交换协议自行协商产生多因子会话密钥，生成的密钥为参与通信的双方独有，端到端加密确保通信内容不被通信参与者之外的任何人获取。采用 SM3 对通信数据进行完整性验证，采用 SM2 对通信内容进行强签名保障通信数据真实可靠，防止遭受中间人攻击。另外，由安全网关将设备的关键参数和系统控制指令等重要信息使用国密算法加密后，结合信息隐藏技术、信道隐写等方法进行隐蔽安全传输，在保证系统正常信息传递的情况下，实现关键信息传输的双重保护。

3.2.3 有效智能分析和挖掘工业互联网安全行为

工业互联网打破传统工业相对封闭可信的生产环境，导致攻击路径大大增加，数据种类和保护需求多样，数据流动方向和路径复杂，因此在多源异构的大数据环境中如何进行智能分析和挖掘有效信息存在很大挑战。本方案在 Spark 框架下，利用机器学习、模糊推理、智能计算等人工智能技术自动获取特定的特征信息，如目标设备的开放端口、提供服务的状态、协议类型等状态信息。基于获取的特征信息，结合行为分析、数据挖掘技术准确挖掘工控设备、协议等相关的脆弱性信息，如系统网络通信中易危端口开放占比及分布、高危漏洞分布地域及类型占比、易忽视的隐患设备运行状态变换、透明的控制过程种类等。安全态势预测在态势理解的数据分析基础上对相关设备或协议的安全等级、攻击手段变化趋势或来源等作出预测分析，全面自动监测系统的安全风险，从全局角度增强对安全威胁的识别。该技术不仅有助于工业互联网的安全身份认证和安全接入，也将有助于工业互联网网络安全态势管理和风险评估。

云桌面在高校计算机实验室的应用研究

郑 舒
（福建省邮电规划设计院有限公司，福建 福州 350001）

摘　要：文章阐述了当今高校计算机实验室的现状与管理的难点、痛点。分析对比虚拟云桌面技术VDI和VOI得出符合实际需求的解决方案，即利用基于VOI技术的搭建虚拟云桌面的应用环境，介绍了该方案的架构设计、硬件分析规划及方案优势。实践证明该方案利用先进的信息技术为教育创新营造有利的环境，既解决数据安全所面临的风险挑战和问题，又满足实验室对本地计算力充分利用的诉求，能有效提升运维人员管理效率，取得良好的实际效果。

关键词：云桌面；云终端；虚拟化；VOI；数据安全

1 项目背景

1.1 建设背景

随着信息技术的飞速发展，教学的需求也越来越多元化，传统的教学方式显然已经不能满足当前教学发展的要求，同时，教学过程中的一系列数据安全问题也严重影响了教学质量和教学进度。如何利用先进的信息技术为教育创新营造有利的环境，解决数据安全所面临的风险挑战和问题，是摆在教育工作者面前的一个崭新且富有挑战的课题。当前高校IT基础设施建设均已投入到不同学院专业的实验、实训、教学等应用。但随着建设规模的不断扩大，管理工作量、管理效率亟待提升。因此云桌面技术开始逐步替代传统PC，越来越多的机房采用桌面云来建设，通过云桌面的模式来简化管理，保证数据安全，提高效率，节省资源，打造绿色低碳的教学环境。

1.2 高校计算机实验室的现状

华侨大学现有计算机实验室因建设年代不同，供货的厂商各异，导致计算机实验室的终端硬件异构严重，性能差异大。同时，专业课程因教材内容，技术进步等各种原因，对教学软件的应用环境，终端硬件的配置要求越来越高，差异化严重，直接导致计算机实验室运维人员工作量逐年增加，工作效率和用户体验下逐年降低，给管理带来困难。

1.3 高校计算机实验室管理的难点和痛点

1）公共课、专业课、考试、技术培训等对实验室应用环境需求各异，导致每个终端上共存着多个系统，数据安全性不佳，且运行速度受到影响。

2）不同的计算机实验室，硬件异构严重，性能差异大，可安装需安装的软件不同，给排课和维护带来困难。

3）当需要使用克隆功能，来完成系统重装、软件升级、漏洞修复等工作时，易受网络环境影响，经常会出现失败终端点，耗费大量的时间和精力。

4） 实验室管理人员人手不足，受学历、年龄结构等因素影响，专业素养技术水平有限。

2 需求分析

教学机房的建设往往需要达成以下目标：

保障教学连续性和数据安全性：教学系统稳健性要得到进一步加强，在学生上机操作时对课件进行处理的时候，数据、系统的安全性会受到来个人自误操作及外部网络的各种威胁，为了防止人为或病毒的攻击，解决方案需要提供一套安全保护措施，充分保证操作系统和数据应用系统的安全，同时要实现在多教室同时授课情况下，对数据和系统的安全性和可靠性的保障措施，灾难应对及恢复策略。

可靠性有保障：各个实验室建设时期不同，网络环境差异大，故对终端设备要求可离线运行，当出现服务器宕机、网络设备故障等情况，可保证业务正常稳定运行，尤其是考试时。

保证教学应用系统及外围设备的兼容性：保证系统在3-5年时间内兼容现有教学应用系统，并能够随时支持新的教学场景切换。

实现高效的集中管理：充分考虑合理的运维费用和较小的运维服务难度，最大限度地降低综合的维护成本，兼顾到不同时期的各项软件应用及硬件设备，利用较小的代价改造这些系统。兼顾到与已有的各种教学管理系统和身份认证系统的兼容性。兼顾多种外接设备，如USB设备，并口设备，串口设备等。

3 桌面云技术与场景应用分析

云桌面是指将计算机的终端系统虚拟化，以达到桌面使用的安全性和灵活性。在任何时间、任何地点，只要有网络，通过任何设备都可以访问属于个人的桌面系统。现如今云桌面虚拟技术在实际应用种发展出两个方向：一类为虚拟桌面基础架构VDI;另一类为虚拟操作系统基础架构简称 VOI。

3.1 业内主流桌面云技术架构介绍

1	VDI	VDI，英文全称 Virtual Desktop Infrastructure，即虚拟桌面基础架构。它不是给每个用户都配置一台运行操作系统的桌面 PC，而是通过在远端的服务器通过桌面虚拟化技术生成多个虚拟机，通过客户端设备访问桌面，用户访问他们的桌面就像是访问传统的本地安装桌面一样。 **VDI 技术的特点是重后端，轻前端，桌面可移动，资源可弹性动态分配**
2	VOI	VOI，英文全称（Virtual OS Infrastructure）架构由服务器来管理操作系统镜像，并下发给终端，操作系统运行于终端本地硬件上。计算任务完全由终端承担，服务器只负责镜像管理、镜像上传下载、以及终端的管理工作。 **VOI 技术特点是重前端，轻后端，系统兼容性完全依赖硬件**

3.2 两种技术架构对比

常见业务需求场景	VDI 架构	VOI 架构
数据不落地本地	✓	✕
资源弹性调度	✓	✕
外设兼容性	✓	✓
开机速度	优	中
网络带宽占用	高	中
断网是否可用，可靠性	✕	✓
原有电脑性能利旧	✕	✓
镜像与应用更新，远程运维	✓	✓
部署服务器数量	多	少

3.3 不同场景的桌面云技术选型探索

序号	教学&实验场景	技术架构选型思路
1	公共机房	主要承担基础计算机信息实验，使用对象为全校学生，不同专业开设的计算机类课程都需要使用公共机房，建设体量大，设备更新周期较长，软件需求多样化、种类繁多，对于硬件配置要求适中。同时，还会承担各类校内、校外的竞赛、比赛。**目前主要会采用 VOI、VDI 技术承载。**
2	经管机房	经管类机房主要分为财务金融软件实践为主体的教学实践机房，基于虚拟仿真的教

		学机房。核心应用为用友、金蝶 ERP 软件、大智慧、同花顺等税务财会类软件，一类是虚拟仿真实训、开放式虚拟仿真实验教学平台软件，针对虚拟仿真有一定图形图像处理需求。 同时，经管机房也会承接一些财经类考试，需要临时部署考试环境。**财务类机房多采用 VDI 技术建设，虚拟仿真实训室根据应用复杂程度采用 VOI 技术建设**
3	3D 类机房	3D 类实验室主要为艺术与传媒学院、建筑类学院、地理信息学院建设，场景偏重图形设计交互显示、偏重渲染计算处理/科学计算，对于终端计算和图形能力要求较高。**目前采用 VDI+gpu 方案方案建设，但由于不同的场景应用存在差异化，因此选型测试会相对复杂。**
4	教师办公	该场景主要是行政办公，主要依托终端完成教学课件的编制，课堂环境的准备，用户数量大但是并发数不高，建设较为分散。**方案选型较为灵活，如果侧重移动办公，则选择 VDI 方案；如果是行政办公，选择 VOI 方案。**
5	电子阅览室	多为学生提供上网及资料查询机，该场景对于资源配备要求相对较低，但对于环境的静音有一定要求，遇到问题能够快速远程恢复。**目前主要采用 VDI 方案进行建设**

4 方案设计

以华侨大学现有的网络环境和实际需求，根据需要，计算机实验室既能满足日常教学需求，又能兼顾对外培训、考试等多种用途于一身的综合性实验室。综合考量之后，选择 VOI 模式，服务器主要功能就是存储数据。在云端根据需求制作各种场景的镜像文件，终端依据需求选择场景镜像文件下载，亦可服务器主动下发，以适应不同的需求。

4.1 方案组成

云桌面方案一般由三部分核心组件组成：云桌面管理服务器、云终端、云桌面软件。

图 1. 管理服务器

如图 1 所示，云桌面管理服务器是为云桌面解决方案而推出的服务器产品，具备易交付、易扩展、易运维三大特点，真正实现了软件定义服务器，让计算和存储资源随需而动。云桌面管理平台能够实现对部署在局域网内的所有节点服务器进行统一监管运维。

图 2. 云终端

如图 2 所示，云终端通过搭载 i5 第十代高性能 Intel 架构处理器，涵盖 8G/16G 内存，512G SSD 存储。可根据实际需求灵活选择，具备卓越的计算性能和超高清视频解码能力，满足高校实训室场景正常教学需求。

云桌面软件是基于操作系统虚拟化技术（集中管理、分布运算、离线可用）。在教学过程中提供便捷、丰富的互动功能的系统，具有屏幕广播、学生演示、语音互动、作业在线提交与在线批改等功能。

4.2 组网设计

华侨大学计算机公共机房，通过采用 VOI 的架构，重点保证使用下发镜像的速率，一般情况下，

在千兆网络下同时下发 60 镜像时，一般需要 10-15 分钟。一旦下发完成，镜像就可以留在本地进行使用。和服务器的网络通信可以使用较低的带宽。如图 3 所示。

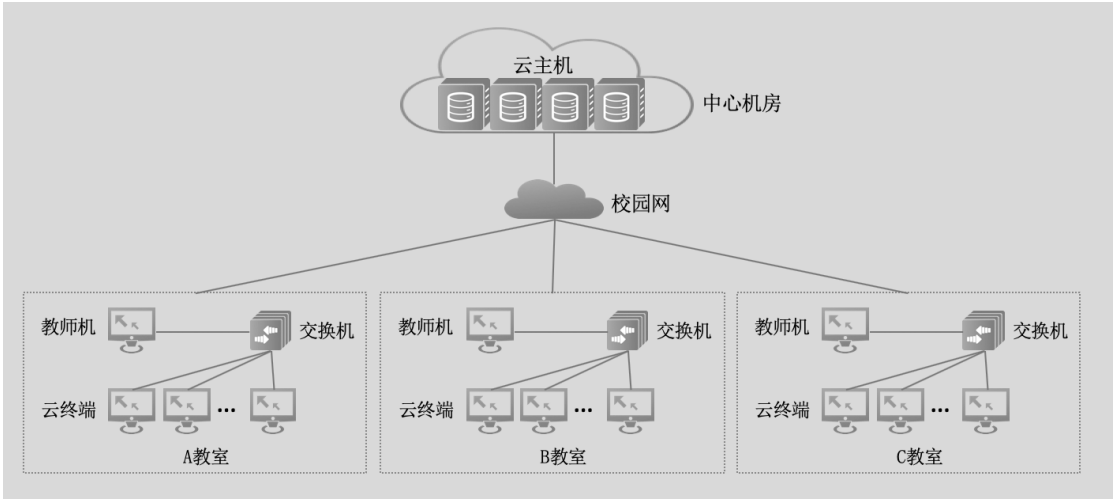


图 3. 组网设计

4.3 存储设计

针对 VOI 类型桌面，无需单独规划服务器端磁盘资源，仅需要考虑终端硬盘可用容量即可，设计如下，如图 4 所示：

图 4. 存储设计

4.4 安全性设计

用户访问的安全性：设置管理员、任课老师、学生身份，仅有管理员身份能够对云主机执行虚拟机生成、修改和删除操作，确保教学镜像等重要数据安全性，任课老师和学生机无法控制云主机，杜绝用户非常规操作的安全漏洞。

策略化的控制：通过系统提供集中的细粒度的策略控制用户的授权访问，针对用户、网络位置、终端环境、应用、云主机等属性决定用户是否能够获得应用的访问。系统允许访问者进行有限制的访问，而不能随意的更改、拷贝信息，更不能将信息带走。

精细化外设管控：严格禁止非授权外设接入适用，对于授权的外设进行精细化管控，确保数据传输的安全性；

用户和设备黑白名单控制：通过管理平台可添加非法用户和非法终端设置，限制非法接入；在信息安全要求高的场合，只允许特定用户从固定地点的终端登录到包含敏感信息的虚拟桌面中，以避免敏感信息遭到泄露。

三权分立，防止越权管理：管理系统中的账号所属角色对应的操作权限进行分离，独立的账号管

理功能模块，管理员根据实际情况分配权限，实现系统管理员、桌面管理员、审计管理员的多级管理；基于集群、主机和桌面池的细粒度权限管理；满足国内涉密安全应用场景。等保标准 BMB17 中明确规定，系统要支持三员分立的管理。即实现系统管理员、安全管理员、安全审计员的权限制衡。如图 5 所示。

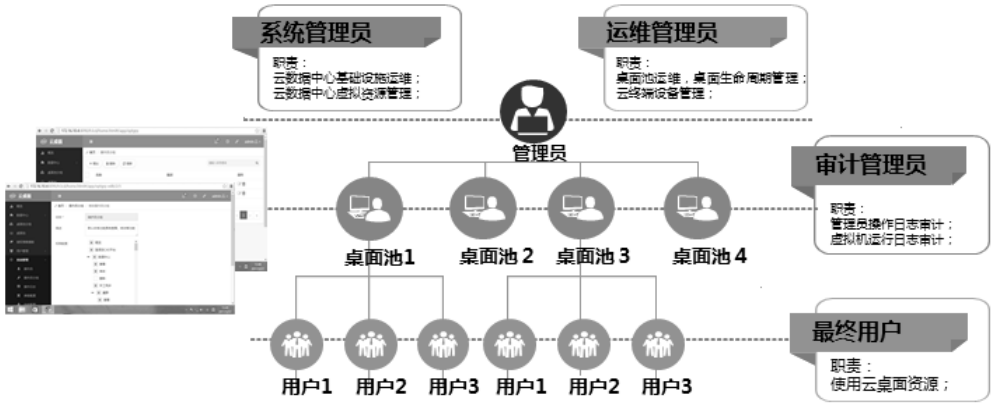


图 5. 安全性设计

5 总结

通过上述的设计及实施，我们发现采用统一的 Web 管理控制台，基于云桌面的实验室可以实现数据的统一安全管理，还能够实现远程运维，对数据中心资源集中管理、统一监控，概览界面直观展示系统健康度、课程/教室/用户总数量、桌面运行情况、资源使用率、异常报警信息等，是未来的实验室升级改造趋势。

5G 定制专网的网络安全部署方案

郑 舒

（福建省邮电规划设计院有限公司，福建省福州市 350003）

摘　要：5G 定制专网的部署要同时满足运营商网络安全运营和客户专网安全使用的需要。先简要介绍 5G 专网部署的几种典型模式，分析 5G 定制专网在接入安全、网络隔离和安全防护几个方面的网络安全需求点，并根据安全需求分析在虚拟专网、混合专网和独立专网几个场景下的网络安全部署方案，为行业客户的 5G 专网部署提供参考。

关键词：5G 专网；网络隔离；安全防护；VPN（虚拟专用网络）

引言

随着 5G 的广泛部署，行业数字化转型升级进入高潮。行业客户的需求多样，应用场景复杂，传统行业专网的传输能力、连接能力、安全性等方面已经无法更好地满足行业客户的需求。5G 定制专网能够凭借 5G 网络大带宽、低时延、高可靠性等特点更契合用户特定的行业需求。然而 5G 专网的安全风险不仅仅是威胁个人通信，更是会影响到行业客户关键部门的敏感系统和运营商大网的安全，因此 5G 定制网的网络安全成为重要关注点。安全保障应该满足业务、等保、上级监管等方面的需求，在网络建设中也应同步规划和建设。

1 5G 专网部署方案

5G 专网是一种局域网，它使用 5G 技术为客户创建专用网络。根据行业客户对 5G 网络覆盖、安全隔离、时延等要求的差异，运营商能为客户提供“切片”“边缘”“独立”的 5G 专网建设方案。

方案一：5G 大网提供专用网络切片

该方案在运营商的 5G 大网为行业客户建设专用的独立核心网切片，通过切片、VPN（虚拟专用网络）、QoS（服务质量）来区分网络和优先级，为客户提供相应行业专属的应用服务，可实现广域跨省、公专协同、业务隔离等功能，既可以灵活部署又可以减少企业的成本。客户接入 5G gNB（5G 基站）访问 5G 大网的 AMF（接入和移动管理功能）、SMF（会话管理功能）和 NRF（网络存储功能）等网元，实现注册和会话建立，选取省级的 UPF（用户平面功能）进行数据承载，通过 UPF 访问企业私网和公网。5G 大网提供专用网络切片的架构如图 1 所示。

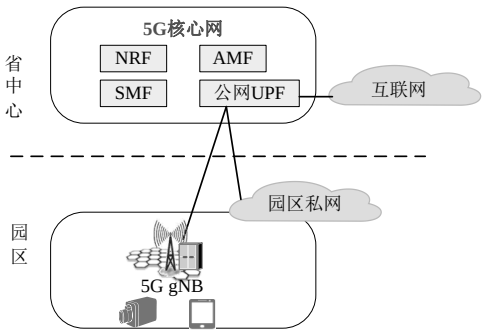


图 15G 大网提供专用网络切片

方案二：边缘节点下沉网元

针对对时延较敏感的客户，可以在临近园区的运营商边缘节点或者在客户机房部署 UPF 和 MEC（边缘计算节点）设备，为用户提供低时延的行业服务。用户可以根据对数据管控的要求选择独享或者与其他企业共享 UPF。因此,该方案适合对时延和数据管控要求相对较高的客户，又能为客户节省部署和维护成本。边缘节点下沉网元的架构如图 2 所示。

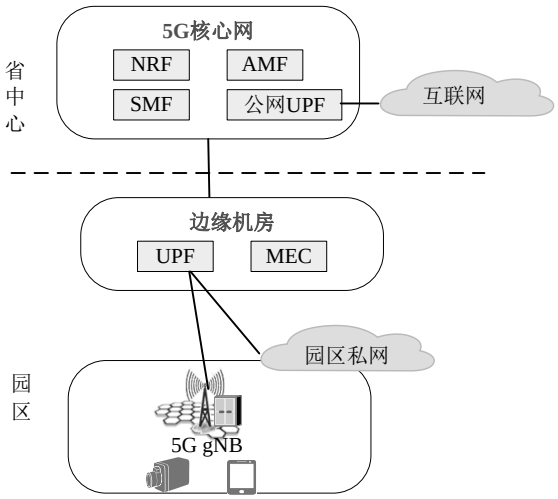


图 2 边缘节点下沉网元

方案三：独立专网

针对要求物理隔离的高安全需求客户，可以为客户建设一张包括 5GC、UPF、MEC 和 gNB 的 5G 专网。物理资源由客户独占，因此能保障企业数据的绝对安全，实现超低时延，缺点是部署成本和运维成本较高。客户园区部署独立专网的架构如图 3 所示。

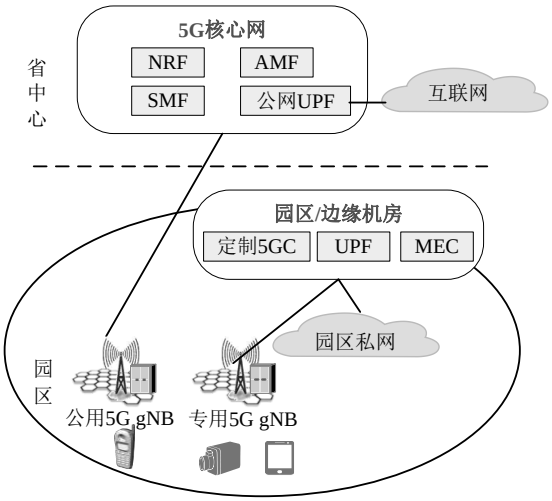


图 3 客户园区部署独立专网

2 5G 专网网络安全需求

5G 专网除了要满足专网用户和大网的网络安全，还需要为不同业务场景提供差异化安全服务。根据全国信息安全标准化技术委员会(TC260)立项的国家标准《信息安全技术边缘计算安全技术要求》，5G 专网的网络安全需求主要包括接入安全、网络隔离、网络安全检测这几个方面。

2.1 接入安全

接入安全是指对接入网络的设备进行身份鉴别，确保访问专网设备的身份合法性，以及用户设备与网络之间、信令面与用户面数据的机密性和完整性。

a) 用户与核心网之间需提供认证服务，空口实施加密策略。

b) 基站设备应支持并开启防泛洪攻击、非法近端登录告警等安全检测功能，并支持告警上报或日志记录，能够在自身设备以及对接网管查看告警信息或日志信息。基站设备应支持 ACL (访问控制策略) 控制功能，能够通过 IP 地址、源端口、目的端口、协议类型等方式过滤 gNB 网口的数据流。

c) 在基站的部署上，可考虑将基站双挂在不同的传输设备下，以便在一个平面的传输资源不可用时，能具备一定的持续服务能力。

2.2 网络隔离

网络隔离是指通过安全路由策略和访问策略，使数据在指定区域流动，避免行业个人信息泄露，也防止 5GC 大网受到非法访问。

由于专网客户防护等级低于运营商要求的安全等级，下沉网元容易被恶意攻击，并且容易通过网络将风险引入 5G 大网，从而严重影响 5G 网络的正常运行。运营商可以将网元分成可信网元和不可信网元。可信网元可以直接与 5G 大网建立接口，流量经过现有的 5G 大网 VPN 承载。如果是不可信的网元，在信令层面上应设置信令网关作为安全网关和信令代理，不可信网元与信令网关建立接口，接入专网的 VPN。如果不具备信令网关，可以在承载网连接下沉 UPF 的设备上进行访问控制，下沉 UPF 到指定 SMF 的报文才能进入专网 VPN。在媒体面上，不可信下沉 UPF 的 N9 接口应接入专网 VPN，并由防火墙实现不可信下沉 UPF 的上行 N9 流量的安全过滤。

5GC 专网的 UPF 与企业私网的对接，建议通过专线等方式实现，UPF 可以通过划分不同的 VLAN 来对接不同的客户私网。

2.3 安全防护

安全防护是通过建设安全配套设备，实时监控

传输流量，及时发现异常并响应，防止专网和大网遭受网络攻击。同时要建立完善的网络应急响应机制，根据流量分析和规则匹配，及时对攻击做出响应，防止攻击流量的破坏，支持常见网络攻击的检测功能。

根据 5G 定制专网安全防护性能要求的不同，安全配套平台可采用本地部署，或者基于集约化的安全能力池进行防护。

对实时性能要求高、流量大，或者与网元紧密结合的防护能力，如防火墙、入侵防范系统、蜜罐等,可与下沉的业务网元一起采用本地部署的方式。

对实时性要求不高的通用安全能力，可采用电信运营商提供的安全能力池进行安全检测和防护，如漏洞扫描、基线核查、数据库审计等。

3 分场景的 5G 专网网络安全部署方案

3.1 场景一：虚拟专网

虚拟专网网络安全方案如图 4 所示。该场景即行业客户采用运营商 5G 大网提供专用核心网切片，专网和公网共享 5G 基站和公用 UPF，网络安全主要是通过 5G 核心网切片配置、网络隔离以及大网的安全能力平台实现。

a）5G 核心网实现对公网业务与专网业务切片间及专网业务切片间的逻辑隔离，或根据业务的重要程度采用独立物理机资源、独立网络设备等实现物理隔离。

b）承载层面应隔离公网业务、专网业务流量，单独划设专网业务的 VPN。

c）该定制网方案选取省网 UPF 进行数据承载，因此采用大网 5G CE 旁挂的防火墙、IPS（入侵防御系统）、恶意程序监测等设备进行访问控制和安全检测。针对安全检测类、分析审计类等对实时性能要求不高的安全能力，如漏洞扫描、基线核查、数据库审计等，调用运营商集约部署的安全能力平台的能力，根据分析数据进行访问控制。

3.2 场景二：混合专网

混合专网网络安全方案如图 5 所示。在该场景下，用户面的 UPF 下沉至边缘节点或者客户机房，同时根据需要同步部署 MEC 等设备。因此，需要考虑下沉网元给 5G 大网带来的风险，以及下沉网元自身存在的风险。

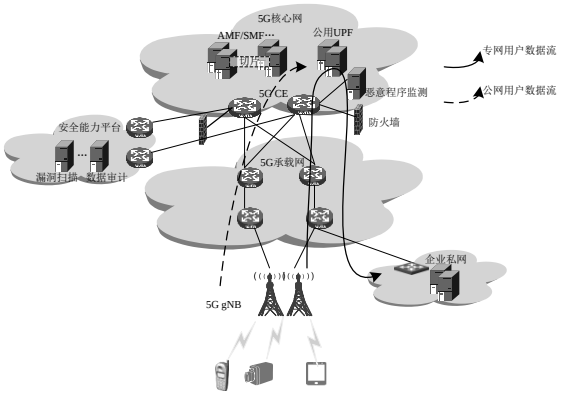


图 4 虚拟专网网络安全方案

a）将下沉网元设置成不可信任网元，与 5G 大网使用的 VPN 进行安全隔离。在承载网连接下沉 UPF 的设备上进行访问控制，限制下沉 UPF 的信令进入 VPN。在条件具备的情况下，在 5G 大网部署信令网关，预防下沉设备的故障或安全风险从外部传导到 5G 大网的内部。

b）在安全防护上，下沉 UPF 所在的边缘节点或者在客户机房部署防火墙、IPS 等设备实现 UPF 与企业私网边界的安全隔离和访问控制。下沉 UPF 在访问大网时，接入专网 VPN，并由旁挂在 5G 大网核心网 CE 的防火墙实现流量的安全过滤。为了节约运营商和客户建设及运维成本，漏洞扫描、基线核查、数据库审计等功能可以调用运营商集约部署的安全能力平台的能力。

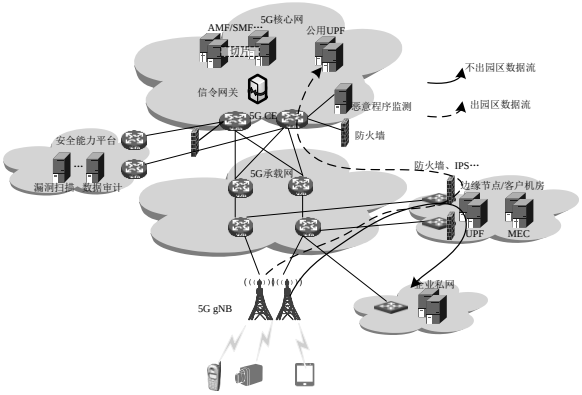


图 5 混合专网网络安全方案

3.3 场景三：独立专网

独立专网网络安全方案如图 6 所示。在该场景

下，客户使用专用基站，采用专用的园区定制 5GC、UPF 设备，按需部署 MEC，保证数据不出园或不出场。因此主要考虑的是独立专网自身的安全风险。

a）私网流量和公网流量根据 NSSAI（网络切片选择辅助信息）实现在基站分流。终端访问私网时，使用定制 NSSAI 实现基站到小型核心网 AMF 控制面的连接，并据此实现控制面和用户面均接入下沉小型核心网。公众应用则统一接入大网 5GC，控制面和用户面均由大网实现。

b）防止企业私网对专网下沉网元的横向攻击。专网与企业私网的边界安全需通过防火墙、IPS 等设备实施访问控制策略，对目标地址是专网设备的流量进行过滤和入侵检测，从而实现独立专网与企业私网边界的安全隔离。

c）根据客户需要，在独立专网内部署基线核查、漏洞扫描等安全防护设备。

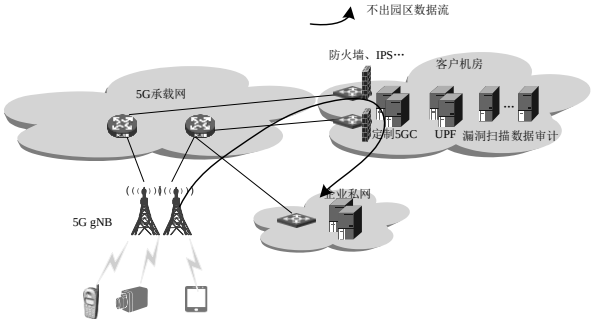


图 6　独立专网网络安全方案

4 结束语

5G 专网业务的终端形态多样，部署环境多元，面临的安全问题越发复杂。在 5G 网络中运行高敏感和高等级安全业务，要求更有效的安全防护体系。随着等保一体机、量子通信、终端安全等技术的逐步引入，5G 专网将为行业客户提供更好的数据隐私性、更安全的隔离度、更灵活的自管理和更高效稳定的连接性能。同时，也能够将运营商的网络安全能力实现开放输出给客户，实现快速获取、灵活自主的客户体验。运营商和设备提供商需要倡导开放合作的理念，深化合作，共同提高 5G 专网的安全保障水平。

参考文献:

[1] 杨红梅，孟楠.5G 时代的网络安全[M].北京：人民邮电出版社，2021.

[2] 饶亮.深入浅出 5G 核心网技术 [M].北京：电子工业出版社, 2021.

郑舒（1985—），女，主要从事固定/移动核心网、信令网、业务网络等方面的规划咨询工作。

构建数据安全长效常态化管控机制解决方案

郑志欢　林宗明　张 恒　雷 佳
（中国移动通信集团福建有限公司，福建 福州　350015）

摘　要：中国移动福建公司高度重视数据安全工作，建立党委数据安全工作责任制，切实贯彻落实党中央、国务院有关决策部署，以习近平新时代中国特色社会主义思想为指导，结合当前数据安全发展面临的新形势新问题，围绕“数据安全管理、数据安全技术、数据安全运营”三个方面，推进数据安全防护体系建设，实现“有制可依、有规可循、有技可施、风险必查”。

关键词：数据安全；数据测绘；行为监控

前言

习近平总书记高度重视数据安全工作，将数据安全上升到国家安全层面，企业内外部形势也面临着较大新的要求和挑战。

1 外有驱动

国家层面出台《数据安全法》及《个人信息保护法》强化数据分类分级管控及个人信息部安全保护要求；

两部委考核对数据安全管理及技术能力方面的提出更高要求，强化数据资产的识别及脱敏、数据流动监测、接口安全管理、安全审计等技术能力；

工信部及集团公司要求全面贯彻落实行业数据安全标准工作，包括基础电信企业分类分级方法、重要数据识别指南、数据安全评估规范。

2 内有需求

资产识别不全，BMO 三域数据量大，结构复杂，敏感信息多、应用场景多，缺少对全量数据资产的自动识别能力；

防护手段单一，数据安全包括数据全生命周期的安全管理，缺少数据脱敏验证、接口自动监测和预警、自动化审计等方面的管控技术；

统一运营困难，各域各自管控，缺少公司级的数据安全能力统一管理视图，实现常态化跟踪运营分析。

因此我们亟需深化数据安全贯标工作，构建数据安全长效常态化管控机制。

福建移动成立了由信息安全管理部牵头，信息技术部、网络部及地市公司等部门配合的贯标专项小组，负责贯彻落实数据安全贯标工作责任，深化公司数据安全贯标工作，其他各部门数据安全贯标专项小组在省公司贯标专项小组的统一领导下，负责开展本部门的数据安全管理工作。

3 目标客户群体

电信、金融、医疗、大型企业

3.1 技术手段

福建移动基于建立健全数据安全管理体系，引入敏感数据检测识别技术、数据访问和操作行为 UEBA 等技术能力，完善了数据安全管理，实现了数据资产分类分级、数据安全访问行为监测、数据资产审计等手段，有效的帮助数据安全管理人员掌握公司数据的分布情况和数据资源脆弱性情况，保障了福建移动数据资产安全，同时获得了 2 篇专利。

3.2 敏感数据检测识别技术

3.2.1 技术能力

资产探测：主动探测网络内存活的数据源信息，帮助企业发现未知或未备案的数据库、大数据组件、文件主机等资产。采集探测数据源的 IP、端口、类型、版本等基础信息，可配置并生效过滤名单。

数据采集：主动连接和访问目标数据源，对数据源内的数据进行切片取样。主动访问 web 对象，对流转数据进行自动爬取，可配置采集参数，支持存储数据源的元数据采集。

数据识别：对采集数据进行处理和解析，利用数据识别引擎，结合敏感识别策略对解析后的数据进行识别匹配，结合的数据分类分级策略，对处理后的数据进行自动化的分类分级标记。支持生效企业基于行业要求和业务现状自定义的分类分级标准和策略。

数据资产测绘：利用分析引擎对数据源、采集数据、识别结果等进行全面分析，实现企业数据资产的全面测绘，形成企业数据资产地图、多维统计分析视图、资产分析报告、资产清单等，帮助企业全面梳理数据资产。

资产管理：协助完成数据源梳理、接入和管理，支持大数据组件、数据库、文件系统、web 业务系统，支持进行业务系统、集群、归属部门、责任人等信息备案和关联，支持连接测试并标记状态。

3.2.2 核心价值

辅助企业完成资产管理和分析，构建数据分析和数据安全管理的坚实基础。

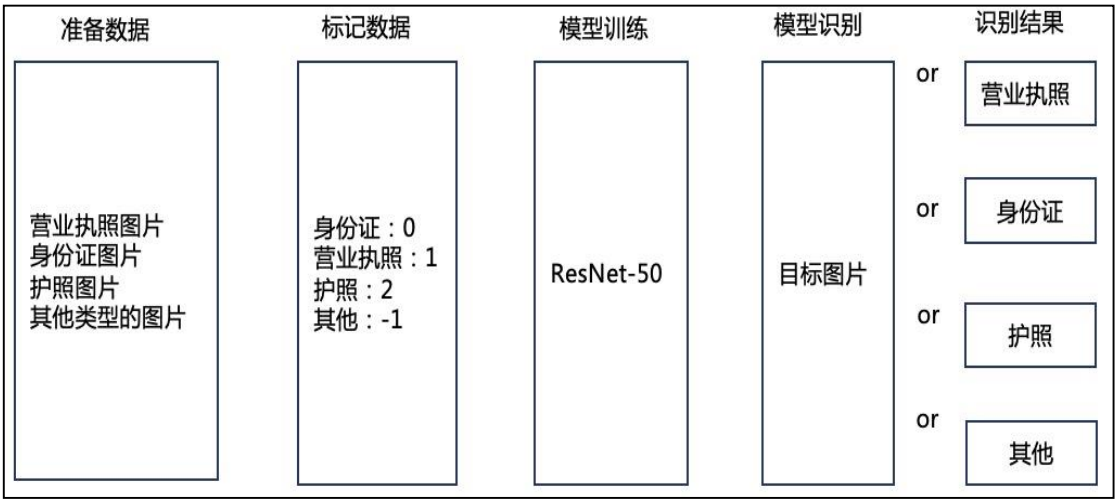
3.2.3 建设价值

实现企业数据资产的系统化梳理、敏感数据或重要数据识别标记、数据分类分级、数据测绘分析，形成数据资产清单、重要数据或敏感数据清单、数据分类分级清单、数据资产测绘分析图表等内容。

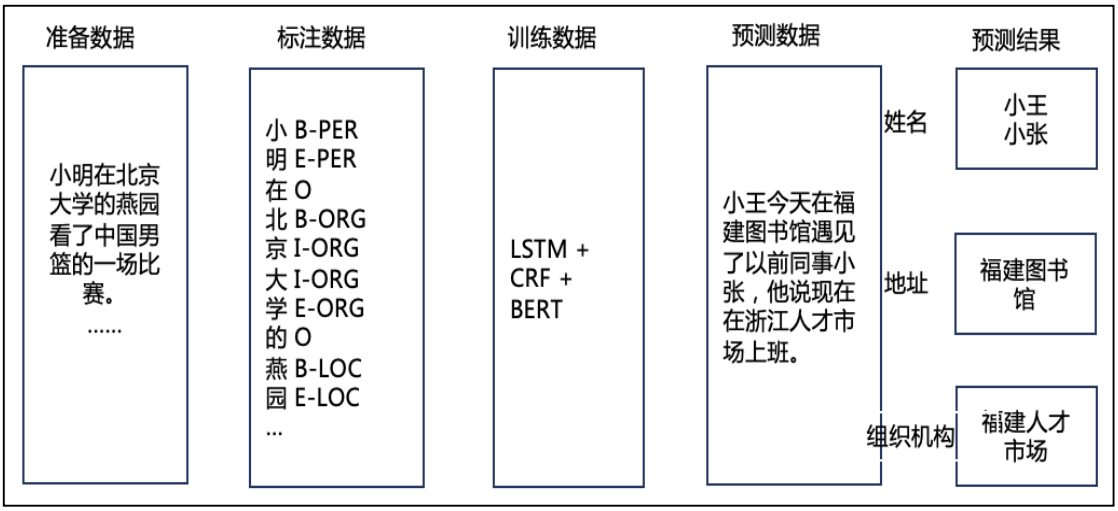
系统通过对企业关键数据进行识别定位，清晰呈现用户隐私数据、业务核心数据等企业关键数据的存储分布情况，为后续针对关键数据的保护和治理工作提供明确的目标和方向。

3.2.4 创新性

通过运用机器学习模型，对某列数据按照自定义的内容字典进行匹配，对于无归纳出数据特征法，使用枚举方式发现的敏感信息字段和其他自定义机器学习算法进行匹配，采用 LSTM 算法用来优化每个句子中前后词语之间的依赖关系，BERT 算法用来优化每一个词语的词向量，基于语义算法明确了数据发现的对象和策略后，对敏感进行识别定位，并对数据进行精准定位，通过人工干预灵活调整发现结果并不断训练过程中，使其更加贴近业务需求，得到更精准智能的识别结果。



基于 ResNet-50 深度学习算法，通过营业执照、身份证、护照等敏感图片特征，根据图片的标签去学习图片中的关键信息，对图片做旋转、加噪点、修改对比度等，同时通过提取关键字对其进行特征匹配及敏感匹配，实现图片的敏感识别。



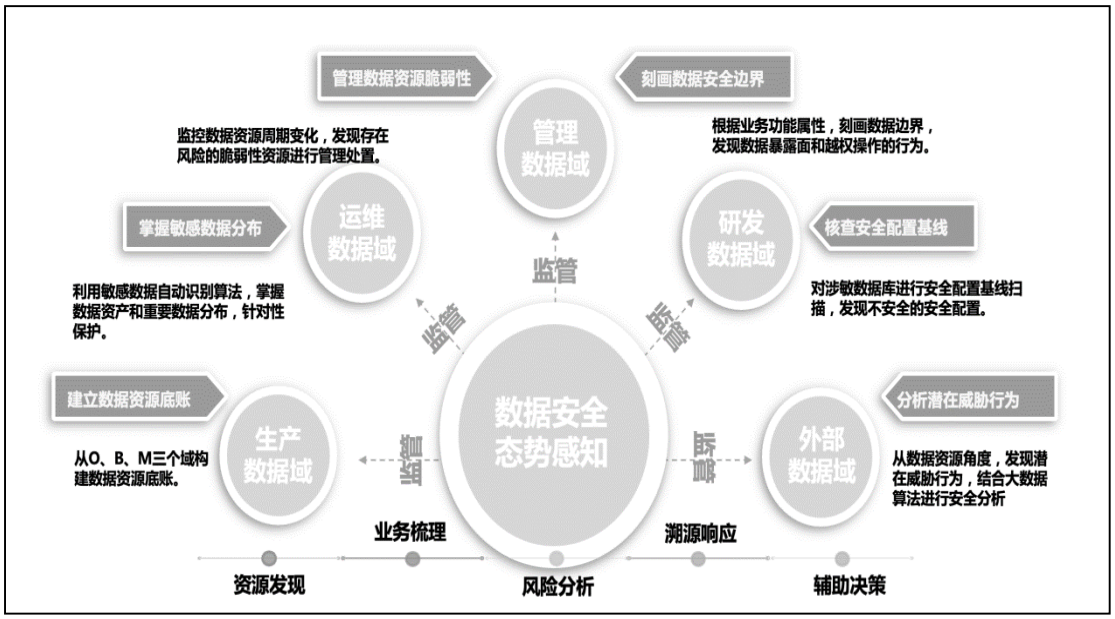
支持通过分发 docker 镜像一键部署，支持采集测绘流程全自动化运行，降低人工运营成本，提升系统易用性。底层探针内置负载均衡模式，支持分布式架构，既能实现轻量化封装，又可承担大型存储集群的资产测绘工作。系统内数据可实现授权共享，可方便实现与其他业务、安全系统联动扩展。

4 数据访问和操作行为 UEBA 能力

4.1 技术能力

通过对业务系统的流量镜像进行实时监测，对敏感数据泄露事件进行监测分析，形成数据泄露事件列表，安全管理人员结合现有的处置手段对数据泄露事件进行协同处置。

敏感数据泄漏监测基于数据源采集的离线数据及实时数据流数据的数据处理能力，大数据分析过程将 ETL 后的流量数据和日志数据存入缓存库、关系型数据库、hive、Hbase、HDFS。利用离线数据对融合 Logistic 方程的 Markov 模型进行训练，完成敏感数据泄露监测的基线学习。利用实时的流式数据对异常事件进行实时审计与告警。



4.2 核心价值

针对载体的脆弱性及合规性进行检查

4.3 建设价值

通过对敏感数据动态风险监测，构建敏感数据泄露监测体系，结合数据监测策略对监测对象的数据安全风险进行动态监测并对策略的准确性进行动态评估从而降低数据泄漏风险。

4.4 创新性

对各个业务系统之间的访问链路关系进行自动化学习。通过训练后的 Markov 模型对业务系统之间的异常数据访问链路关系进行识别。利用正态分布算法，对正常访问接口数据流量范围、接口访问数据范围、接口访问链路等信息进行智能学习。无需人工参与即可完成对外接口日常访问行为的分析，建立合理的日常访问基线。

融合 Logistic 方程与 Markov 模型及其他核心算法，对用户日常操作行为进行画像分析，通过对模型训练后可形成用户日常行为基线。包括用户日常访问时间、用户日常操作习惯、用户日常访问敏感数据范围及数据量级等。通过与实时流量对比，即可识别用户异常操作行为。

与用户行为分析类似，载体行为分析主要侧重对数据载体的访问及操作。包括载体日常访问源 IP 范围、载体日常访问频次等信息。同时本创新点与合规系统保持联动，针对载体的脆弱性及合规性进行检查。

5 统一安全管理能力（4A 系统）

5.1 技术能力

集中帐号管理：为用户提供统一集中的帐号管理，支持管理的资源包括主流的操作系统、网络设备和应用系统；不仅能够实现被管理资源帐号的创建、删除及同步等帐号管理生命周期所包含的基本功能，而且也可以通过平台进行帐号密码策略，密码强度、生存周期的设定。

集中认证管理：可以根据用户应用的实际需要，为用户提供不同强度的认证方式，提供具有双因子认证方式的高强度认证（一次性口令、数字证书、动态口令），而且还能够集成现有其它如生物特征等新型的认证方式。不仅可以实现用户认证的统一管理，并且能够为用户提供统一的认证门户，实现企业信息资源访问的单点登录。

集中权限管理：可以对用户的资源访问权限进行集中控制。它既可以实现对 B/S、C/S 应用系统资源的访问权限控制，也可以实现对数据库、主机及网络设备的操作的权限控制，资源控制类型既包括 B/S 的 URL、C/S 的功能模块，也包括数据库的数据、记录及主机、网络设备的操作命令、IP 地址及端口。

集中审计管理：将用户所有的操作日志集中记录管理和分析，不仅可以对用户行为进行监控，并且可以通过集中的审计数据进行数据挖掘，以便于事后的安全事故责任的认定。

5.2 核心价值

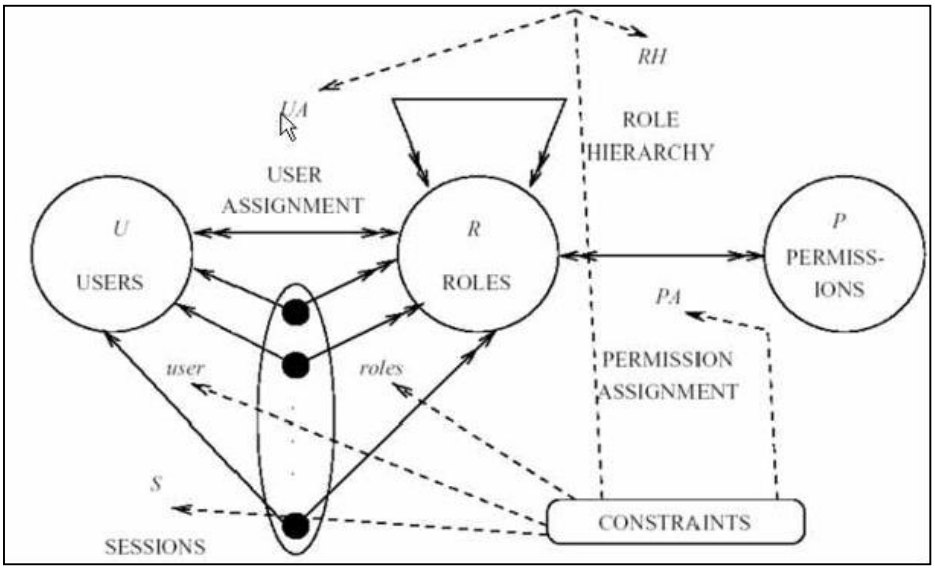
可以为企业提供全局的帐号管理视图，有效控制随意创建帐号、僵尸帐号等带来的安全管理问题，实现基于角色、菜单的细粒度应用资源授权管理使得企业可以清晰地梳理资源与人员间的关系， 及时发现不符合权限的人员、帐号共用等问题，基于业务场景的精确审计分析和预警。

5.3 建设价值

依托 4A 金库、个人文件夹、零下载功能，从数据的访问获取、传输、使用、销毁各个阶段实现全生命周期的管控，防止数据外泄。

5.4 创新性

基于角色的访问控制在访问控制中引入了角色的概念，把对资源的访问权限分配给相应的角色，根据用户在组织内所承担的角色进行访问授权与控制。一个用户可以承担不同的角色，从而实现授权的灵活性。只要某用户属于某个角色那么他就具备这个角色的所有操作许可，即该角色所拥有的权限。用户与角色是多对多的关系。



6 效益分析

敏感数据检测识别技术建设使用后，通过自动化采集存量数据资产样例数据、元数据等，基于敏感数据分析引擎实现对敏感数据的自动识别、分析及分类分级标记，减少了敏感数据人工梳理工作量，提高了敏感梳理的工作效率。

通过输出敏感数据识别结果输出到静态脱敏、动态脱敏、金库管控及审计系统，辅助实现敏感数据的安去管控，协助防止数据泄漏造成损失数百万元。

数据访问和操作行为 UEBA 能力建设运营后，采用智能敏感数据泄露监测的方式大幅度的降低了人工监测数据泄露与数据安全风险审核的工作，节省了人力成本提升了整体数据安全风险管控能力，对重点安全事件的自动预警、自动派发、一键处置的联动管理能力，初步实现网络安全问题的闭环管理。

4A 系统系统建设使用后，企业范围内统一的权限授权和相关金库管控，使企业对所属信息资产的安全管控能力达到新的高度，减少管理人员工作量 1/2。帮助福建移动建立起敏感信息金库防护体系，使安全配置隐患可在第一时间被发现，并得到迅速解决，保障 IT 系统正常运行，最大程度减少企业的运营风险。

绿盟数据安全解决方案

王　玉
（绿盟科技集团股份有限公司，福建 福州 350001）

摘　要：随着云计算、大数据、物联网、人工智能等新技术的发展，网络边界被不断打破，敏捷创新、安全合规驱动快速转型，社会和企业都在面临数字化的转型带来的数据安全风险。数据安全已经与关键信息基础设施一并成为影响社会安全的关键因素。本项目通过绿盟基于数据安全防护体系的解决方案，实现安全管理与技术手段相结合。

关键字：数据安全 顶层设计 安全治理

0 数据安全建设体系

在数据安全建设体系上，我们提出“一个中心，四个领域，五个阶段“的顶层设计。一个中心是指以数据安全防护为中心。四个领域是指的数据安全建设的四个领域：组织建设、制度流程、技术工具和人员能力。五个阶段是指的数据安全建设的五个阶段：业务梳理，分级分类，策略制定，技术管控，优化改进。在数据安全建设体系中，组织建设、制度流程，技术工具，人员能力四个领域都需要同步开展建设工作，组织层面，决策层、管理层、执行层必须在数据安全建设领域达成一致，数据安全建设工作必须得到组织高层的支持。组织高层在数据安全领域的战略目标应该能够被管理层和执行层实现。我们日常所说的“三分技术七分管理”也好“七分技术三分管理”也罢，都是在表明，管理是技术的运营依据、技术是管理的落地保障。所以两者要相辅相成，缺一不可。数据安全建设体系中，我们借鉴了 Gartner 的数据安全治理框架，定义了数据安全建设的五个阶段，形成了绿盟的数据安全方法论。总结起来就是“知”、“识”、“控”、“察”、“行”五个字。“知”是指制定数据规范、定义敏感数据；“识”是指数据分类分级、发现敏感数据；“控”是指知道安全策略，控制敏感数据；“察”是指敏感数据安全监察、数据行为追踪溯源；“行”是数据安全持续运营。

1 数据安全建设流程

数据安全解决方案建设流程主要分为六个阶段。

业务数据梳理。在组织与制度设计方面，传统网络安全均由 IT 部门负责，随着数据治理工作的深入开展，业务部门要深入参与数据资产梳理以及分级分类工作之中，因为只有业务部门是最了解数据价值与重要性的。因此需要自上而下形成高层牵头，横跨业务部门与安全部门的组织架构。由信息安全管理团队和数据业务管理团队共同商讨建立数据安全制度流程体系。制定好制度体系应该以文档化的方式进行落地管理。从最高级的方针战略，到最细节的表格日志，都应该由不同层级的团队负责进行文档化的落地，并严格执行。在相应的业务组织与管理制度指导下，企业才能更好的开展后续建设工作。

定义与识别敏感数据。开展数据安全治理的第一步就是：定义什么是敏感数据，基于业务特点进行数据的识别、数据分类、数据分级。数据分类分级的准确清晰，是后续数据保护的基础。由于数据类型不同，对企业影响不同，我们建议根据《中华人民共和国网络安全法》要求对个人信息和重要数据分开进行评估与定级，再按照就高不就低的原则对数据条目进行整体定级。

数据全生存周期安全风险评估。完成敏感数据

分类分级后，就要到风险识别的步骤：发现哪里有敏感数据，并对敏感数据进行梳理与风险评估。敏感数据发现与数据风险评估的工作要结合人工服务和专业工具共同完成。数据安全风险评估可以从数据的生存周期角度逐个考虑，这里引用国标GB/T37988-2019《信息安全技术 数据安全能力成熟度模型》DSMM 架构图中的数据生存周期安全的步骤：数据采集安全、数据传输安全、数据存储安全、数据处理安全、数据交换安全、数据销毁安全。绿盟科技研发了敏感数据发现与风险评估系统，可以实现智能数据分类分级、全网数据资产测绘、实时数据流转测绘、大数据平台风险扫描的能力。结合上绿盟的数据安全评估服务，从数据生存周期各个阶段评估数据安全风险，应该可以帮助您解决很大部分的敏感数据发现与风险评估问题。

数据安全纵深防护。针对数据安全的风险，应以数据为中心，向外对业务、网络、设备、用户采取“零信任”的态度，既然每个环节都不可信，那么管控手段就要覆盖全部环节，任意环节失信后都能实现熔断保护。用户侧、终端侧、网络侧、业务侧，以及数据中心，都要做好安全防护措施，外向内防攻击防入侵防篡改，内向外防滥用防伪造防泄露。最关键的是，要对全部纵深防护环节进行整体控制，实现环境感知，可信控制和全面审计。整合多层次的纵深防御，及时发现问题，及时阻止安全问题。总之我们的防护宗旨是认证好人并允许其通过，识别坏人并阻断其访问。

敏感数据监察分析。敏感数据监察分析、发现安全问题与异常事件。可以考虑从用户、资产和数据的行为模式出发，利用 5W1H 分析模型来进行敏感数据行为分析，基于行为模式发现数据异常事件。也就是我们常说的 UEBA。 基于历史的可信访问行为提取访问规则，利用各类算法进行行为聚类，形成可划分的访问行为簇并可视化呈现。通过这种图谱分析与可视化展示让管理者对于敏感数据访问情况，由一无所知转变为可视可管。

优化改进与持续运营。当我们具备“知识控察”的能力后，不代表我们的数据是安全的。业务是在变的，数据也是在变的。因此我们的安全也是要不断变化的。为了应对变化，我们在“知识控察”的基础上提出了“行”，这是一个动词，代表着对数据安全的优化改进与持续运营。在大的层面，合规要求指导安全策略的设置，安全策略支撑合规治理要求的落地，二者相辅相成，配合上持续优化改进运营的“知识控察行”体系，实现持续自适应的数据安全防护能力。

2 数据安全建设难点

为更好的实现组织信息安全防护，满足业务发展和监管合规要求，各单位正在全面积极开展信息系统安全保护类项目的建设，防护能力已经基本覆盖了网络、终端、数据和业务系统。但是，随着安全建设的进行，安全性和易用性之间的矛盾也日益突出，同时对于数据安全仍然缺乏针对性的措施来配合防护，主要表现在以下方面：首先是核心数据难以梳理。数据存放在电脑、手机、笔记本、业务系统、数据库、存储中，大量的结构化、半结构化、非结构化数据难于辨识，无法判断出哪些数据是重要的，保护难以下手；无法对数据标准定义，无法有效对数据分类分级；无法明确核心级别的数据在公司的整体分布情况；缺乏对核心数据生命周期的风险评估，数据生命周期安全现状不明确。其次是管理制度难以落地。业务环境复杂，管理制度不够切合业务流程，制度推行缓慢；企业现有制度流程不够完善，安全现状合规程度低，难以有效解读法律法规，实现企业安全合规；没有有效的安全工具帮助安全管理制度落地，保障企业安全合规。另外，内部泄密难以管控。虽然网络、业务系统、终端、存储多个维度都在管理，但是防护难以实现数据安全管理紧耦合；各单位间存在文件交换的需要，外发的文件存在一次性复制、终身拥有、无限泄密风险；数据以刻录光盘和内网邮件等范式传输，文件接收人可以在任意环境使用文件，存在文件失控扩散风险。在日常工作中，不可避免的要通过打印、刻录、IM 传输、网络发送、邮件等方式发送敏感数据，无法有效的管理和控制这些行为，防止重要文件的非受控扩散。最后，泄密风险难以追查。重要的核心数据缺乏整个流通通路的监控审计，难以追责到人；数据可以通过移动存储介质轻易的从公司电脑拿走，内部专用介质无法有效限制使用范围，

介质滥用无法追溯审计；安全事件发生后，如果没有一套完善的行为审计系统，仍然无法进行及时告警响应、准确定位事件源头，给企事业单位带来极大的困扰和严重的信息安全隐患。

一个成熟的数据安全体系应当做到技术、管理相结合，管理范围全覆盖，事前防御、事中管控、事后审计和整体风险态势可分析。同时结合企业当前数据安全管理现状，需要解决分散控制、过度防护等带来的防护效果和业务影响问题。实现企业复杂业务场景下的全面数据安全防护，形成数据安全的事前、事中、事后、全程把控的技术体系，将技术体系作为工具辅助安全管理。

将数据安全管理做到看得见、管得着。因此我们需要在信息泄露事件发生之前能够预警、提前防范；在数据泄露事件发生时能够智能的实时主动防护；在信息泄露事件发生之后做到及时告警、审计、快速响应和定位；在安全管理上做到掌握企业整体数据安全态势。即在预警、防护、响应和分析四个层次上进行安全管理。

3 数据安全防护体系解决方案

随着国家电网信息化建设的深入，绝大多数信息是以电子文件和数据的形式被管理和存储，信息化使国家电网信息的生产、存储、获取、共享和传播更加便利，然而，与此同时，非常重要需要保护的内部人事任命文件、国家绝密文件等，伴随网络和办公设备的自由使用却因为缺乏有效的信息安全管理机制造成保密信息泄漏，给国家电网信息安全带来更多的威胁，随着国家电网湖南省电力公司内部信息安全问题日益严重，信息系统的数据安全成为信息科技建设的重中之重。

本项目通过绿盟基于数据安全防护体系的解决方案，实现安全管理与技术手段相结合的目的，在省公司信息内网部署部署终端 DLP 管理平台，在内网关键终端部署 agent，通过事前主动防御、事中检测响应、事后追踪溯源、全程态势感知的整体防护理念，帮助企业建立完善的整体数据安全管理体系。实现数据泄露防得住、数据风险看得见、安全管理好落地的整体效果。可针对运营商、金融、政府、交通等多个行业的客户。

数据分级管控：通过“密级标识”、“透明加密”及“权限管控”模块，实现根据文档价值等级自动进行分级管控。对高价值等级文档进行严格的权限管控，防止泄密的同时，限制其在企业内部的使用范围；对于一般价值等级文档进行加密管控，防止文档泄密对企业造成损失；对于可公开的文档，选择不对其进行管控，满足业务使用需求，最大化平衡安全与效率。

数据使用权限管控：通过权限设定，对企业重要核心数据资产进行保护，通过权限管理可控制重要资产的使用访问和权限，可对文档进行细粒度的权限访问控制，包括只读、编辑、打印等。文档授权后，授权用户只能在授予权限范围内使用文档，非授权用户无法打开使用，从而实现对企业重要数据进行有效的安全防护。

介质管控：对移动存储介质使用权限进行管控，防止内部数据被拷贝泄露到外部非法存储介质中。通过注册内部专用介质，实现内部数据在存储介质中的安全保护，即使专用介质丢失也无法在企业外部查看到介质中存储的数据。

数据外发管理：特殊情况下，需要将内部核心数据外发给合作伙伴，通过文档权限管理，保护外发的文档安全，防止核心机密文档被合作伙伴有意或无意泄露。

4 解决方案创新性

本项目的创新性在于通过“事前主动防御，事中响应检测，事后追踪溯源”的数据安全技术体系，充分保障企业数据和业务安全，显著降低企业生产运行过程中面临的数据安全风险威胁，提升企业网络安全整体保障水平。

5 解决方案意义价值

同时，有一定的意义和价值。满足国家法律法规相关要求。本项目落实了国家网络安全战略层面及国家电网网络安全工作要求，帮助企业挖掘数据安全管理中的风险点与薄弱点，建立合规有效的数据安全管理体系。通过此项目的落地，为整个行业提供科学有效的样板效应，在数据价值体现的同时保障数据的合规安全。

实现可视化的数据风险态势感知。本项目通过

可视化的数据安全管理能力，使企业数据安全防护工作更高效，更清晰，利用人工智能和大数据分析技术，让数据安全数据统计及分析更准确，掌握企业整体数据风险态势，通过图谱分析与可视化展示让管理者对于敏感数据访问使用流转等情况，由一无所知转变为可视可管。随着改革开放的不断深化和信息化的快速发展，企业在发展中产生了大量的数据资产。这些拥有自主产权的数据资产已成为企业竞争力的根本基础，那么如何保护企业核心数据资产安全，如何维持自己的研发创新力，已成为关乎企业生存发展的重要问题。同时，党和政府高度重视数据应用及安全保障相关工作。《中华人民共和国国民经济和社会发展第十三个五年规划纲要》明确指出，要“加快推动数据资源共享开放和开发应用”；要“加强数据资源安全保护”，“保障安全高效可信应用”。

由于电子信息传播途径广、传播速度快、拷贝/复制操作无痕迹及易携带等特性，普通的网络边界防护措施已无法有效控制有内部人员参与的泄密事件的发生。根据国家安全机关的统计，现在泄密事件超过 80%的发生均与内部人员有关，暴露出了企业内部监管手段的薄弱以及安全管理体系的缺乏。

一旦发生机密数据或者信息资产泄密，带来的损失和深远影响，将无可计量。所以如何保证企业内网中的核心数据资产安全是关系到企业生存发展乃至国家安全的重要问题，因此绿盟诚挚希望能够协助您建立完善的数据泄露防护体系。

快快游戏盾 SDK 防御系统解决方案

林思弘　黄斌寿　杨雪云　刘 杰
（厦门快快网络科技有限公司 厦门 ）

摘　要：DDoS(分布式拒绝服务)是目前最突出的攻击。抗 DDoS 攻击在网络安全起到了至关重要的作用,可以帮助服务器在遭遇 DDoS 攻击时,免受卡顿,被访问甚至直接崩溃的时候帮助清洗异常流量。防御 DDoS 攻击不仅具有挑战性，而且具有战略意义。本解决方案旨在帮助企业解决 DDoS 攻击的威胁，减小其带来的损失。

关键词：安全；防御；溯源攻击；DDOS；CC

0 背景

全球 DDoS 攻击持续爆发,攻击峰值不断创记录。DDoS 攻击的行业分布方面，游戏行业成为最大受害行业，占据近 40%的攻击。在游戏逐渐成为大众休闲娱乐的主流方式背后,硬件与开发技术相继发展成熟,超高的热度和丰厚的市场利润回报为整个游戏行业的持续发展埋下了巨大隐患,层出不穷的非法网络攻击行为愈演愈烈。除此之外，网络服务（占比 4%）和企业门户（占比 2%）也是攻击占比较高的行业。

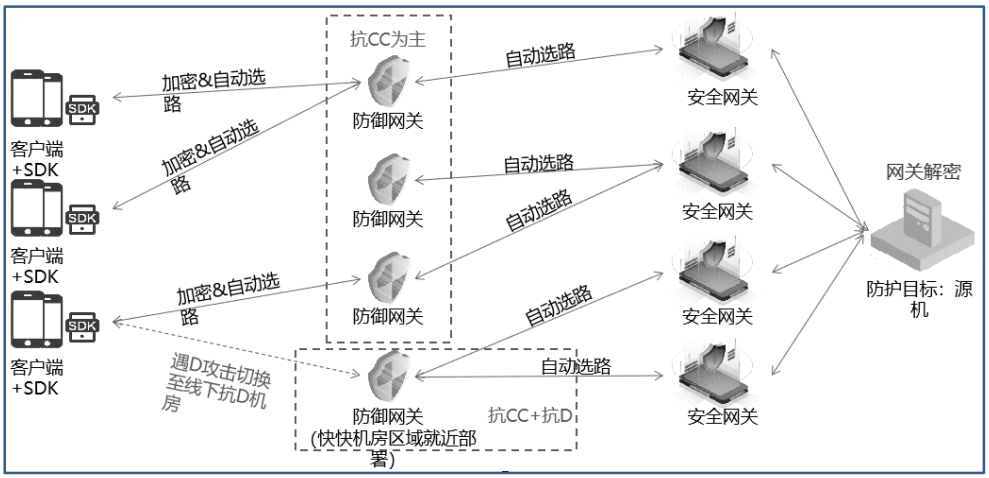
纵观整个网络安全环境,安全已经不再是单独的个人或者游戏厂商能够应对的。如何应对层出不穷的网络攻击已经成为游戏行业首要解决的问题。游戏市场亟需一个有效稳定的安全解决方案。

1 解决方案介绍：

游戏盾 SDK 是针对游戏行业面对 DDoS、CC 攻击推出的针对性的网络安全解决方案。游戏盾 SDK 提供内含 windows、安卓、IOS 版本的 SDK 下载，通过接入 SDK，由 SDK 接管所有的通信流量，进行调度和加密传输，满足抗 D、防 C、流量加密等业务需求，为用户提供优质的网络环境，游戏极速畅通无阻,相对于高防 IP,SDK 效果更好，且更智能化。

游戏盾 SDK 在风险治理方式、算法技术上全面革新,帮助游戏行业用户用更低的成本缓解超大流量攻击和 CC 攻击,解决以往的攻防框架中资源不对等的问题。

1.1 架构图：



1.2 服务内容：

智能加速、断线重连：智能规划优质网络传输路线，游戏加速不断连。

防护 DDoS：攻击通过分布式的抗 D 节点，同时基于 SDK 端流量数据的灵活调度策略，有效将黑客攻击进行拆分和调度，使之隔离。

无惧 CC 攻击：游戏安全网关配置 SDK 建立加密通信隧道，仅放行经过 SDK 和游戏安全网关鉴权的流量，彻底解决 TCP 协议层的 CC 攻击。

溯源攻击：基于用户画像，将进入黑名单库的终端设备视为可疑的攻击源。

个性化定制：可根据客户的需求，定制不同的功能，如专线加速， 模拟器行为分析等。

1.3 应用场景：

1、游戏发行受同行攻击及黑客勒索，游戏稳定性出现问题，服务器宕机及影响游戏正常运行，需求自动防御，抵御攻击。

2、游戏源站 IP 暴露，引发针对性攻击，需求 SDK 高强度加密通信。

3、无针对玩家分配合理网络线路，影响部分玩家游戏使用体验，需求可智能规划优质网络传输路线，确保游戏加速不断连。

4、分布式计算平台遇到同行攻击，影响其正常访问，需求外部攻击防御自动响应。

5、各种互联网应用及平台，防御攻击等。

1.4 案例分享：

助力某类传奇手游抵御大型 DDOS 流量攻击

1.4.1 客户需求：

1、高防御

2、CC 防御能力

3、时延 100ms 以下，网络波动无影响

1.4.2 我司提供的解决方案:

1.解决大流量攻击问题

采用分布式高防御节点作为防御网关，可抗 1.5Tbps 流量攻击，如有更高需求，还可增加云堤清洗。

2.解决 CC 攻击问题

SDK 端与安全网关建立加密通信隧道，仅放行经过 SDK 和游戏安全网关鉴权的流量，彻底解决 TCP 协议层的 CC 攻击。

解决网络波动问题

基于 SDK 的网络链路诊断功能，智能选取优质网络传输路线，保证游戏时延最低，并创建断线重连机制，哪怕玩家本地 4G/wifi 网络异常，也不会导致游戏掉线。

1.4.3 防御效果：

1、当客户的部分防御网关节点遭遇攻击峰值达到 650Gbps(300G TCP + 350G UDP)的 DDOS 攻击，CC 新建连接并发量达到 30w 时，无玩家反馈异常。

2、CC 攻击期间，查看防御网关节点 cpu 使用率为 25−30%，并未达到警戒线 60%。

CC 攻击类型为模拟传奇包攻击，由于未通过 SDK 加密协议，所有的 cc 流量直接被防御网关拦截，未透到客户的源服务器。

2 目标客户群体

2.1 服务行业

互联网、手游、直播、金融等行业或大型企业。

2.2 客户画像

1、客户及其同行经常受到攻击、带宽量不大、对稳定性有较高的追求。

2、认为高防 IP 防护力度不够，因为高防 ip 通常只有 300G 的防护，且容易存在误杀和误报的情况。

3 解决方案拟解决的问题

游戏盾通过嵌入 SDK，针对游戏行业面对的 DDoS、CC 攻击推出的针对性的网络安全解决方案，除了能针对大型 DDoS 攻击（T 级别）进行有效防御外，还具备彻底解决游戏行业特有的 TCP 协议的 CC 攻击问题能力，防护成本更低，效果更好！安全有效，为企业业务安全护航！

4 创新性、先进性说明

通过专业及多年的线上实战对抗经验积累，将黑灰产的攻击手法进行透彻研究，用数据和自研特有算法实现链路加密、智能调度及弹性资源对抗；

具体如下 ：

1、隐藏本地资源，防止本地破解、抓包获取攻击目标；

2、智能风控调度，动态隔离风险用户；

3、高防资源兜底，资源对抗，硬性资源储备兜底对抗。

4、链路加密和可信通信：自研高强度动态加密算法实现一机一密，一链一密。加密隧道保证数据传输安全，黑客无法抓包获取域名、URL 、API 等业务相关内容数据，无法做任何伪造和重放攻击请求，防御抓包、重放、伪造攻击。

5 推广前景

5.1 经营风险与对策分析

5.5.1 项目组织管理及条件保障

1、公司软硬件设施保障

快快总部位于厦门，软件园二期有租赁办公楼，配有内部机房，以及数百台开发用服务器和数百台开发用 PC，有成熟的软件研发技术流程和配合流程，同时有完善的选，育，用，留一条龙人才服务。

公司拥有自主知识产权 86 项，其中发明专利授权 2 项，申请发明 9 项，实用新型授权 14 项，软件著作权 50 项，商标注册证书 12 项。公司运营多个数据中心，管理服务器十万台，拥有总带宽出口 20T+，其中济南数据中心是济南联通核心资源已由快快网络独家签约，厦门海峡数据中心是福建省规模最大的 IDC 基地，安溪数据中心向东南沿海地区辐射的重要战略基地，是华东规模最大的高可用、高安全的数据中心。

2、公司管理

公司自成立以来一直注重产品的技术研发和人才管理，公司通过设立企业技术中心，加大与客户同步开发的力度等措施，有效适应市场需求的变化。同时提出了采用“产、学、研”科技创新模式，以自主创新为动力，以机制创新为保障，以管理创新为支撑，全面提升企业科技创新能力，力求通过实施科技创新，解决和突破其在产业化进程中面临的诸多技术难题和发展瓶颈，进一步发展和完善公司产业化经营模式，全面增强企业核心竞争力。

（1）标准化管理，取得管理体系认证证书

快快网络目前已通过 ISO9000 质量管理体系认证、ISO20000 开发服务管理体系认证、ISO27001 信息安全管理体系的认证、CCRC 信息安全服务资质认证等。

（2）产学研用协作情况

快快网络目前已经与厦门理工学院深入开展产学研合作，委托学校研究《基于智能博弈的网络安全风险预警技术》项目，主要研究大数据分析技术，从海量告警日志中分析挖掘出高危的攻击特征，实现智能性网络入侵检测技术，基于博弈论与人工智能理论，研发基于智能博弈的网络安全风险预警技术。

双方合作进行技术攻关，并约定以甲方名义申请网络安全相关技术发明专利 1 项，合作发表 SCI 核心期刊论文 1 篇，指导甲方申报科技进步奖 1 项或合作申报科技进步奖。通过产学研合作，进行网络安全技术创新，合作申报科技进步奖。

5.1.2 经营风险分析和对策

无论游戏盾 SDK 所服务的行业，还是所处的抗 DDOS 细分领域，既有乐观的市场需求增长率，也有竞争不断加剧的趋势。网络安全市场吸引了国内外大量网络安全知名企业涌入，知名网络安全企业基于自身对产品的需求而研发产品技术，并投入市场，使得市场竞争日渐激烈。

1、技术层面对策

项目技术基于快快网络 8 年开发，依托（云）主机安全产品、Web 应用防火墙、网络安全解决方案，构筑预测、防护、检测、响应的动态自适应安全模型，从数据的感知、认知到预知，意图建设全网免疫系统，为公有云、私有云、混合云等各种云环境下的用户提供全面保障，拥有多项安全领域核心专利，以迎合未来市场多方面需求。

2、市场层面对策

项目已在工业制造、金融、游戏等多个行业得到成功应用，现有用户一万多家，充分运营云、大数据、Al 等技术，构建（云）主机安全平台、Web 应用防火墙、网络安全解决方案，为用户打造集"预测、防御、监控和相应"一体的安全闭环，市场前景良好。同时公司的服务遍及全国 26 个省市，拥

有先进的服务体系、安全专家团队，为用户提供7*24 小时快速相应服务。核心团队多年深耕云安全行业，技术人员来自网络安全、主机安全等领域，获得众多信息安全认证。

3、游戏盾 SDK 技术风险分析及对策

（1）风险分析

首先，本系统接受大量的外部数据，存在信息系统自身安全的脆弱性问题，系统应对错误的数据和结构不合理的数据进行识别，拒绝接受错误数据和结构不合理数据。其次，本项目涉及到众多大数据相关技术的统筹应用，风险可能来自于项目技术环境如开发环境、技术、软件及硬件、外围系统集成、产品升级等。另外，在不同种类的软、硬件设备，同种设备的不同版本之间，由不同设备构成的不同系统之间，以及同种系统在不同的设置条件下，导致系统可能存在各自不同的安全漏洞问题。

（2）对策

公司在云安全的关键技术上已有多年的研究经验，并且已经研发出了成熟的云安全产品。公司将通过提升系统的健壮性，及时升级系统和补丁软件，建立完善的安全管理制度等方式解决系统技术风险。同时本次项目可充分利用产学研用多方合作优势，将理论支持、技术研发、场景需求与实践应用有机结合。通过召集产学研用多方专家交流研讨，对于项目中可能出现的风险进行技术风险预测，对于意外风险进行应急预案制定，对于已经发生的风险进行及时的解决方案研制，对于超规模技术风险确保有效外部支持。

4、法律风险和对策

公司主营数据中心业务、云计算和云安全业务，是国家工业四基、六基中的重点发展行业，也是工业强国战略中的锻长板、补短板业务；公司依法经营，分别在公司组织架构、行政办公、工作考勤、销售绩效、财务管理、合同管理、知识产权等各个方面制定了 30 多个管理制度；常年聘请法律顾问，对公司的内外事务进行合法性审查。

5.1.3 公司管理模式和激励机制

公司采用信息化系统支撑管理，业务上云，数字化赋能主营业务。钉钉系统管理着人事考勤和 OA 审批系统，自研的数据中心资源管理系统管理数据中心的云计算、云安全业务，自研的销售 CRM 客户管理系统实现售前、售中、售后各种任务管理。研发中心根据反馈到的各种问题升级迭代程序，使系统不断强壮。

设立员工持股平台，制定股权激励计划。公司于 2021 年 3 月请公司法律顾问制定了《2021 年股权激励计划》，主要股东拿出 20%股权，400 万股，注入股权激励池，设立员工持股平台的有限合伙企业，对公司高管和技术核心人员开展股权激励计划，第一批 5%，100 万股，经考核合格的 19 人已顺利实施股权激励计划。

5.1.4 财务数据

快快网络基于云计算云安全对数字经济的促进作用以及过去多年以来的发展状况，未来营业收入目标实现年增长率 30%以上，实现净利润高增长，实现研发费用投入高增长，实现弥补历年亏损后可多交税。2022 年到 2024 年的财务数据（单位万元）预测如下：

年度	营业收入	营业成本	所得税	净利润
2019	6295.9	4629.76	0	19.91
2020	11530.93	8753.98	0	257.77
2021	15205.99	11631.98	0	531.66
2022	22000	14300	125	975
2023	30000	20000	500	1660
2024	42000	29400	750	2736

6 证明材料

6.1 应用证明

序号	文件名称	类型
1.	华为鲲鹏技术认证书	资质证书
2.	华为技术认证书	资质证书
3.	厦门悦游网络科技有限公司	应用证明
4.	厦门护卫云信息技术有限公司	应用证明

6.2 专利

序号	专利号	名称
1.	ZL 2018 1 0992080.1	一种安全的云备份系统及方法
2.	ZL 2021 1 0743385.0	一种基于人工智能的工业智能制造产品品质全流程控制方法及控制云平台
3.	ZL 2021 2 2851436.X	一种云计算服务器保护设备
4.	ZL 2021 2 2851477.9	一种云计算机主机防爆安全存放柜

6.3 软著

序号	软件著作权登记证书名称	登记号	首次发表日期
1.	快快游戏盾 SDK 防御系统 V1.0	2021SR1810658	20211120
2.	快快游戏盾管理系统 V1.0	2019SR1382707	20191122
3.	快快域名白名单对接系统 1.0	2017SR330916	20161229
4.	快快云 CDN 自动化管理系统 V1.0	2017SR330342	20160624
5.	快快云 CMS 管理系统 V1.0	2017SR331130	20160919
6.	快快云弹性计算软件 V1.0	2018SR894169	20180226
7.	快快云盾安全防御软件 V1.0	2018SR894190	20181024
8.	快快云防御体系邮箱监控报警系统[简称：云防监控]V1.0	2017SR331119	20160405

6.4 其他

序号	材料名称	类型	备注
1.	国家高新技术企业	国家	GR202035100472
2.	厦门市科技小巨人领军企业	厦门市	20190701
3.	厦门市专精特新企业	厦门市	202008
4.	厦门人才企业榜科创板潜力企业 TOP20	厦门市	2019 年度
5.	2021 年度福建省数字经济领域瞪羚创新企业	福建省	202104
6.	国家反诈中心第二届优秀警企联盟企业	公安部	2021
7.	福建省软件和信息服务 50 强	福建省	20211124
8.	厦门市重点软件和信息技术服务企业	厦门市	20210916
9.	2017 年度优秀云计算服务商	IDC	201712
10.	2019 中国云安全领域杰出方案提供商	中国产业互联网协会	20190418
11.	2020 中国网络信息安全创新产品	中国信息协会	202007
12.	林思弘 2019 最佳青年榜样	CFS 中国财经峰会	201907
13.	参与国标 20204692-T-604 号起草证明	国家标准	正在公示中
14.	ISO9001 证书	质量管理体系认证	2019-202212

		证书	
15.	IS020000 证书	信息技术服务管理体系认证证书	201912-202212
16.	ISO27001 证书	信息安全管理体系认证证书	201912-202212
17.	信息系统安全等级保护备案证明	三级	202109
18.	CCRC 证书	信息安全服务资质认证证书	202202-202502
19.	计算机信息系统安全专用产销售许可证（web 应用防火墙）	公安部	0503212473
20.	科技成果转化项目证书	厦门市科技局	20200821
21.	快卫士安全专用产品销售许可证	公安部	0404220381
22.	山西省市场监督管理局感谢信		2021.1.21
23.	重庆市铜梁区公安局感谢信		2021.1.26
24.	华为技术认证书	游戏盾 V1.0 与华为云鲲鹏云服务完成解决方案联合测试	202109-202408
25.	鲲鹏技术认证书	游戏盾 V1.0 与华为云鲲鹏云服务完成兼容性测试验证	202103-202402
26.	鲲鹏技术认证书	HUAWEI ENABLED 证书及认证徽标的使用权	202106-202405

AI 无障碍反电信诈骗技术“彩印”：反诈正名、亲情联防、大数据精准反诈

王 乐　李 鹏　黄嘉崴　王 欢　许益峰　梁玉麒
（咪咕动漫有限公司 福建 厦门）

摘　要：在电信诈骗犯罪手段逐步升级，个人信息泄露、正向预警被误拒、反向提醒不及时等问题并存的大背景下，为构建"全警劝阻,全民联防"新格局,咪咕动漫有限公司以 AI 大数据技术为驱动，结合运营商通信能力，按照“正向反诈劝阻身份标识，反向疑似诈骗联防提醒”的组合拳策略，以期逐步构筑覆盖全国的电信诈骗防护网。

关键词：双向反电信诈骗；AI；大数据；联防提醒

1 目标客户群体：覆盖面广，痛点刚需

1.1 反诈名片目标客群：主叫公务人员

反诈名片主要针对工信部反诈中心、国家反诈中心、各级公安机关、政府机关、社区网格员等企事业单位人员。这类目标用户群体存在这样的核心痛点场景：当前公众对“陌生来电”几乎零容忍，陌生来电接听率仅约 5%−10%，高拒接率直接影响了政府机构、企事业单位等的工作效率和人工成本。

1.2 亲情彩印目标客群：被叫弱势群体

1.2.1 老人及其子女

我国人口老龄化问题日益凸显。2022 年 1 月，国家统计局发布数据表明，我国 65 岁及以上人口已破两亿（占 14.2%。[1]），预计 2033 年突破 3 亿[2]，部分学者认为这意味着我国已进入深度老龄化社会，今后高龄、独居空巢老人将进一步增多。老人身体机能较青壮时期大幅下降，判断力也同步下降；空巢，精神空虚，缺少陪伴；老人手中有部分财产。故，家有老人的子女们核心痛点场景：不法分子利用金融、信息通讯技术的广泛应用，频频实施电信诈骗活动，对老年人的财产安全和身心健康造成严重伤害。

1.2.2 学生、留守孩子及其父母

处于“象牙塔”与现实社会交界地带、涉世未深的学生群体（包括大学生、大专生），缺乏社会经验，安全防范意识薄弱，对网络诈骗的花样套路认知不足，手里又拿着父母的血汗钱。同时，学生群体的信息被网络应用软件不当收集和使用，甚至被售卖给不法分子。

而且，诈骗已从城市向农村蔓延，长期生活在农村的留守孩子由于性情淳朴、消息闭塞，对各种诈骗套路基本难有识别和防范的能力，农村诈骗成功率远高于城市。

1.2.3 初入职场的青年

当前刚刚开始工作的青年，正是中国开始计划生育后的新生代，绝大部分被父母当成掌中宝，生活技能、社会阅历普遍欠缺，自我保护、自我防范意识薄弱；且初入职场消费需求上涨，但缺乏稳定或充沛的经济来源，同时网络上消费诱惑多，青年想要赚快钱，容易落入诈骗陷阱。

1.2.4 家庭主妇

随着女性地位的不断提高，可支配经济的自由度也越来越高，消费需要包括但不限于医美健康、穿搭美妆、教育提升、亲子育儿、公益爱心等，即便妇女接受教育的程度也在逐步提升，但是电信诈骗为利所驱，防不胜防。

1.3 12381 目标客群：大数据反向精准追踪全国

潜在受害者

12381反诈劝阻平台主要是针对全国范围的潜在受害用户，如“曾经”或“多次”与诈骗分子存在电话及短信联络的用户。涉诈预警劝阻系统主要是基于公安机关提供的涉案号码，利用大数据、人工智能等技术自动识别涉案号码联系过的电话号码，并及时向这些电话号码发送预警劝阻短信，提醒用户警惕潜在受骗风险。

2 解决方案拟解决的问题：正反两向提醒阻力大

诈骗防护正努力构建一个可信通话的体系。防止电信诈骗有个核心拟解决的问题：在于通话过程信息不对称，无法识别陌生来电号码。

2.1 正向预警劝阻号码被误拒：耽误预警时机

近年各级公安机关对正在遭受电信网络诈骗的群众进行预警劝阻，收效显著，但在实际反诈工作开展过程中，群众往往把公安机关的预警电话误认为是诈骗或骚扰电话而拒接，影响了预警劝阻成功率。

2.2 反向疑似诈骗联防提醒难：高门槛、低信度、强延后、弱联防

反电信诈骗应用使用门槛过高。国务院今年印发的《“十四五”国家老龄事业发展和养老服务体系规划》强调，实施积极应对人口老龄化国家战略。报告指出：适老型产业发展进入“攻坚期”，但行政资源、信息化、人力资源以及场景应用等方面仍缺乏明确的规制与标准，应用场景载体与老年人生活、使用习惯的匹配度有待加强[3]。

反电信诈骗信息库信度低。多数应用信息来源不够权威，号段更新也不及时，容易出现误判。

反电信诈骗提醒延后。电信诈骗一般会在通话过程中获取目标核心信息，尤其金融诈骗，一般确认信息后，被诈骗对象就难以反悔逆转局面。多数阻隔提醒手段，往往在通话结束后被关注，甚至在通话期间被拦截。

反电信诈骗提醒没有联防力度。关键亲属（如老人的子女、学生/青年的家长、孩子父母、家庭主妇的丈夫）不知情，无法即时劝阻。

3 创新性及先进性

3.1 彩印四大能力创新：双向提醒、零门槛、不被拦截、可联防

1） 来电正反身份即时同步：帮助用户了解陌生来电身份，经由核心号库识别来电号码身份，通过彩印平台多手段触达，为用户提供“正反”向提醒服务：正向提示黄页号码，为用户展示工信部、公安机关等的正向来电，即推送反诈名片；反向提示风险来电，提醒用户防范可能的诈骗骚扰侵害。

2） 零门槛：零流量、零时延、零安装。目前互联网行业推出的具有通话亲情联防的 app 需要户主及成员均进行安装，使用过程中需保持数据网络畅通，且提示推送可能存在延时；彩印均不需要。

3） 不被拦截。彩印触发基于通话信令，覆盖通话界面，不被拦截。

4） 联防提醒。为深入贯彻落实习近平总书记关于打击治理电信网络诈骗犯罪工作重要指示精神，咪咕动漫依托中国移动和彩印业务平台推出亲情彩印，户主一人开通，全家共享反诈提醒：当亲情彩印成员接收到被彩印平台识别为疑似骚扰诈骗电话或境外来电时，系统将通过下发彩印的方式通知户主及成员注意辨别，防范诈骗风险；通过双重提醒达到家庭联防效果，即做到双重保护。中国电信、中国联通等友商目前主打个人通话防护，暂未推出亲情联防类型的产品，亲情彩印属运营商首创。2021年亲情彩印入选中国移动“安心行动”重点产品，截止目前已有超 500 万用户开通，为用户推送提醒超 2 亿次。

3.2 技术方案五个先进：号源权威、识别精准、联防推送、触发信令、信息智能

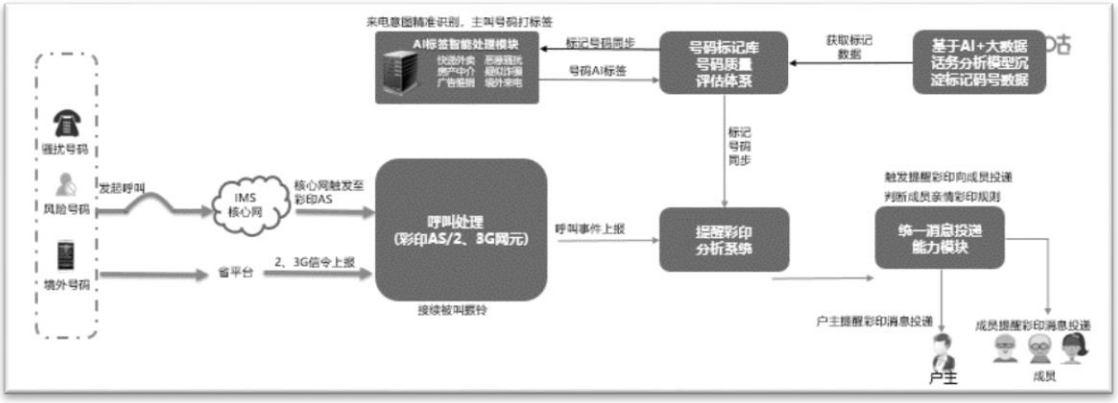
1）号码标记库模块：权威号码源获取与分析。负责获取外部标记号码源、接收号码 AI 标签，运用中国移动海量的通话数据，通过大数据分析能力对号码、呼叫数据进行分析建模、形成号库模型，为提醒彩印及亲情彩印业务提供标记数据。

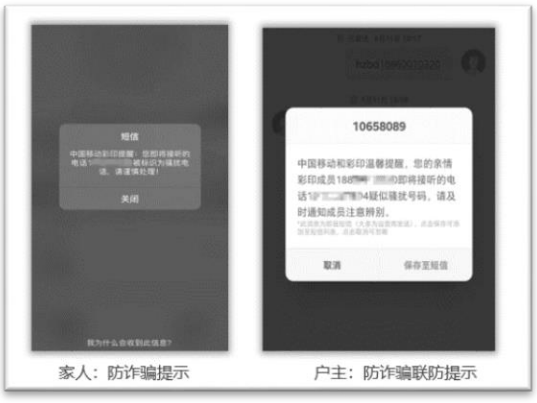
2）AI 标签智能识别处理：主叫号码标签，来电意图精准识别。接收呼叫事件上报，根据呼叫行为特征分析，通过 AI 技术对用户数据及标记号码进行识别分析处理，生成行业号码、境外来电及诈骗高危号码等特定场景自分析号码标签数据。

3）统一消息投递能力模块：联防关系网规则判定。负责接收彩印 AS/2、3G 网元呼叫信息上报并执行彩印业务触发投递，包括接收呼叫事件、实施用户过滤和鉴权、组装彩印内容等，并根据投递策略，执行彩印投递，以 USSD、闪信的方式将彩印内容投递给手机终端。

4）彩印 AS/ 2、3G 网元 ：触发信令，保障防诈信息不被拦截。彩印 AS/2、3G 网元接入 IMS 核心网、SCOM 等，负责对接各省 IMS 核心网 S-CSCF 网元，接收 VoLTE 及 2、3G 彩印用户信令触发，实现呼叫控制，并转发呼叫信息至统一消息投递系统进行业务触发，在振铃中即弹屏提醒。

5）提醒彩印分析系统：智能形成提醒信息。提醒彩印分析系统根据彩印中央管理平台提供的每日呼叫清单以及多个第三方号码库提供的号码，生成标记号码库。当彩印用户接听或拨打标记电话时，统一消息投递系统通过标记号码库对另一方号码进行信息识别，并将识别到的信息以 USSD 或闪信的方式告知彩印用户。



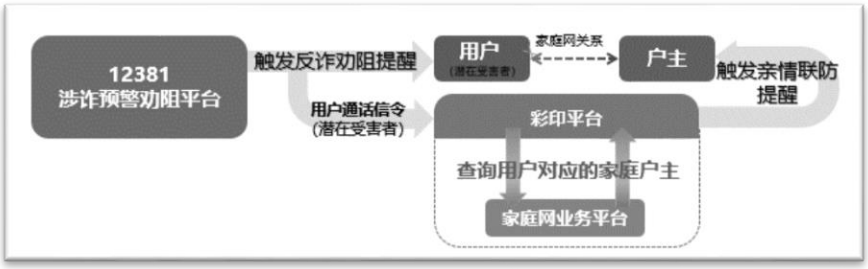


4 大省推广案例及试点效果

4.1 北京 12381 反诈劝阻亲情联防试点：警企合作更护民

工信部反诈中心联合国家反诈中心，组织运营商等推出了“反诈名片”，对各级公安机关的反诈预警劝阻电话号码进行标记和来电提醒，帮助群众有效甄别电话来源，进一步提升预警电话的权威性和及时性。

2021 年 10 月起，彩印平台接收 12381 平台提供的反诈劝阻信息，通过与北京移动合作查询亲情网家庭关系，匹配受害者的家庭户主，同时为保证时效性，数据匹配后满足在反诈劝阻下发 24 小时内，彩印平台则向户主发送反诈亲情联防提醒，截止 2022 年 6 月，共计发送超 4500 条提醒信息，联防效果良好。



4.2 与广西亲情网业务融合：深耕家庭防护市场

为推动亲情彩印业务及中国移动家庭网业务发展，共同深耕家庭网络与数据安全市场，提升用户粘性，将亲情彩印与广西亲情网进行业务融合，以点带面，带动用户规范发展，实现收入模式创新。通过三个阶段开展大数据精准营销，通过掌厅、短厅、网台、基站等渠道，营销曝光率达 3000 万+次。

4.3 湖南家庭共享安防权益：纳入亲情彩印提醒

为发展家庭融合业务，拓宽新业务使用人群，结合本地实际发展需要，湖南移动针对家庭共享会员产品升级权益，将亲情彩印纳入家庭共享权益中，同时推出优惠订购政策，促进双方业务共同发展。

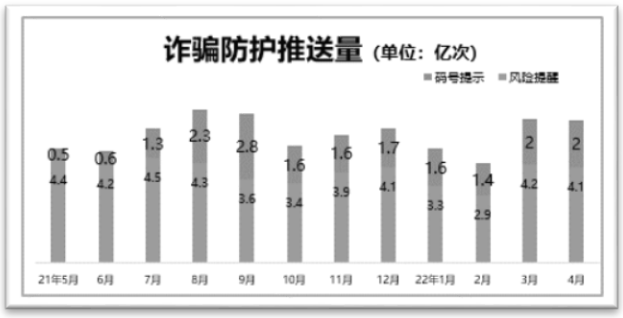
优惠政策：1)开通亲情彩印并首次办理家庭共享会员的用户，年合约有效期内持续使用移动通信网络服务，即可享受前 6 个月 1 元/月优惠（标准资费 3 元/月）；2)已办理亲情彩印的用户，首次开通家庭共享会员，即可长期享受 6 折优惠办理服务。

5 彩印业务发展数据情况

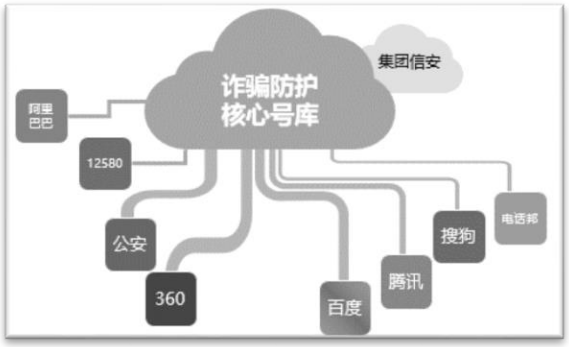
5.1 诈骗防护服务概况

诈骗防护于 2017 年上线防诈骗来电号码提示服务，服务依托彩印平台，截止目前用户规模达 1.04 亿，号码库累计超 2700 万规模，月均风险提醒约 3.9 亿次、码号提示约 1.6 亿次。

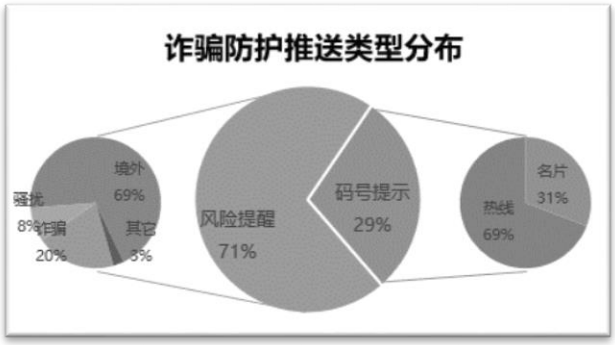
5.2 诈骗防护推送量



5.3 诈骗防护核心号库



5.4 诈骗防护推送类型分布



参考文献:

[1]王萍萍：人口总量保持增长 城镇化水平稳步提升 _中国经济网——国家经济门户 (ce.cn)

[2]中国人口中长期趋势：2050 年中国总人口 13 亿左右 | 生育_新浪财经_新浪网 (sina.com.cn)

[2]全力推动养老服务高质量发展_中国人大网 (npc.gov.cn)

多院区集团化医院网络智能化运维建设思考

黄国强
（福建医科大学附属第一医院， 福建 福州 350003）

0 背景

随着云计算、人工智能、SDN、区块链、物联网、5G 等技术的发展，推动着企业的数字化转型，业务模式不断创新，网络规模快速增长，对网络的可用性要求持续提高，任何一次网络服务中断都有可能对业务造成极大影响。当业务发生故障时，网络运维团队必须快速、准确、有效地定位是否是网络故障导致，若是则需要快速找到故障根因，及时解决问题并保障业务稳定可靠运行；若不是也需要给业务部门解释和澄清网络运行状况。网络运维团队面临着巨大的挑战。

传统运维只看设备的网络通断和零散的指标，指标采集周期长，问题发现往往依赖于用户、业务人员投诉等事件驱动，不能及时发现问题，导致网络运维团队经常处于被动的状态；

依赖运维经验和手工操作的传统故障排查方式效率低下，面对规模庞大业务应用复杂的网络，一旦出现问题，故障原因难以快速定位和解决；

网络新技术发展，规模和复杂度的增加，对运维人员的技术能力要求大大提升，但运维团队却面临人力成本控制，无法相应地增加人力，矛盾日益凸显。应对上述挑战，突破困局，必须利用大数据和人工智能等技术来赋能网络运维，让网络运维从自动化向智能化发展，才能在最短时间内发现问题、定位故障根因和解决问题，降低 MTTR(平均故障修复时间)，保障业务稳定运行。

福建医科大学附属第一医院(以下简称“附一”) 创建于 1937 年，是集医疗、教学、科研于一体的大型综合性三级甲等医院。医院综合实力雄厚，是福建省高水平医院，并被国家发改委、国家卫健委确定为首批全国疑难病症诊治能力提升工程项目医院和中国罕见病协作网福建省牵头单位。医院有茶亭院区、滨海院区、奥体院区、闽南医院，总编制床位 4500 张，已形成“一院多区、一体多翼、协同发展”的办医格局。

1 运维建设目标

基于福建医科大学附属第一医院的定位及现状，早在 2000 年就制定了信息化发展战略，并十年如一日持续夯实发展，积累了丰富的经验。2018 年 4 月，国务院办公厅关于促进“互联网+医疗健康”发展的意见（下文简称《意见》):

- ✓ 健全“互联网+医疗健康”服务体系；
- ✓ 完善“互联网+医疗健康”支撑体系，提升医疗机构基础设施保障能力，及时制订完善相关配套政策；
- ✓ 加强行业监管和安全保障。

结合《意见》的指导，我院的信息化建设目标也快速聚焦到区域信息平台、互联互通、大数据、医学人工智能、互联网医院等五个方面，在 2026 年前希望通过互联网+医疗模式，借助信息手段为医院战略需求的达成提供更大的助力。

信息中心肩负了全院 3 大院区的 IT 网络建设及运维工作，运维要求高，运维人力紧张是长期存在的挑战，伴随着医院信息化战略的推进，IT 网络运维工作也急需一次智能化的升级变革。

2 转型思路

围绕运维建设目标，福建医科大学医院信息中心在医院 IT 网络运维领域提出了五大转型思路：

- ✓ 运维工作从 IT 基础架构视角向业务视角转型

传统的运维工作更聚焦于 IT 基础架构视角的运维，无论是硬件、软件，还是应用的管理工具，

基本都是竖井式建设，形成了信息孤岛。在医院快速发展的背景下，亟需通过业务视角打通信息壁垒，完善运维数据信息融合，从业务视角审视整体运维工作的执行情况。

医院信息中心长期以来通过信息化手段为医院其他业务部门提供业务支撑，保障医院的医疗等业务工作顺利开展，而随着医院业务发展的要求和信息化水平的发展，信息中心作为医院信息主管部门也要完成角色的转变，从单纯的业务支撑提升到业务使用上来。通过更加主动地参与到业务流程中，利用信息化手段让各业务系统的数据流转起来，并提供业务流转和优化的参考。

✓ 运维平台从零散化向集约化转换

以往医院的信息化运维通过多种手段或者工具完成，实际使用中需要维护多套系统和不同的工具，不同手段各有其特点，对于运维人员就需要熟悉更多的系统和工具，这本身就增加了运维人员的工作量，分散了宝贵的人员精力。而一体化运维作为运维发展的趋势，可以有效解决这一难题，通过一体化的运维平台，降低了人员熟悉异构平台的难度。

✓ 服务管理体系向标准化、流程化重构

在长期的医院建设发展过程中，信息部门形成了一套满足自身需求的服务管理体 系，但伴随着福建医科大学附属第一医院日益发展的医院规模，原有信息化部门承担的服务管理系统也面临着越来越大的压力，运维人员需要更加稳健和便捷的管理方式对运维服务过程进行有效管理。

✓ 引入运维自动化、智能化、可视化等先进技术能力

通过技术手段，提升运维的自动化、智能化程度，把运维人员从繁杂、重复的工作中解放出来，使有限的人力更加聚焦在医院信息化发展的主航道。 IT 运维工作的转型本身就是一个需要循序渐进的过程，特别是在福建医科大学附属第一医院这样一 个承担了如此多重要民生健康工作职责的机构中，运维转型工作更加需要谨慎的开展。

经过充分的调研考量，结合医院整体工作的规划和现状，融合运维平台建设围绕以下五个方面开展：

✓ 统一运维建设：重构或更新零散的运维工具集。建设 IT 资源全面监、管、控的统一运维平台；

✓ IT 服务管理建设：对信息中心 IT 服务管理流程、运维管理制度进行规范化建设， 落地各项管理流程电子化流转；

✓ IT 资产管理建设：实现对 IT 设备的生命周期、维修、维护、维保等情况进行管理，随时了解各项 IT 资产状态；

✓ 运维自动化建设：简单重复的日常巡检工作实现自动化，解放部分人力资源；

✓ 运维可视化建设：运维管理各项工作可视化呈现，数据中心现状、业务系统健康度、各类运维报表、手机 APP 远程管控等。

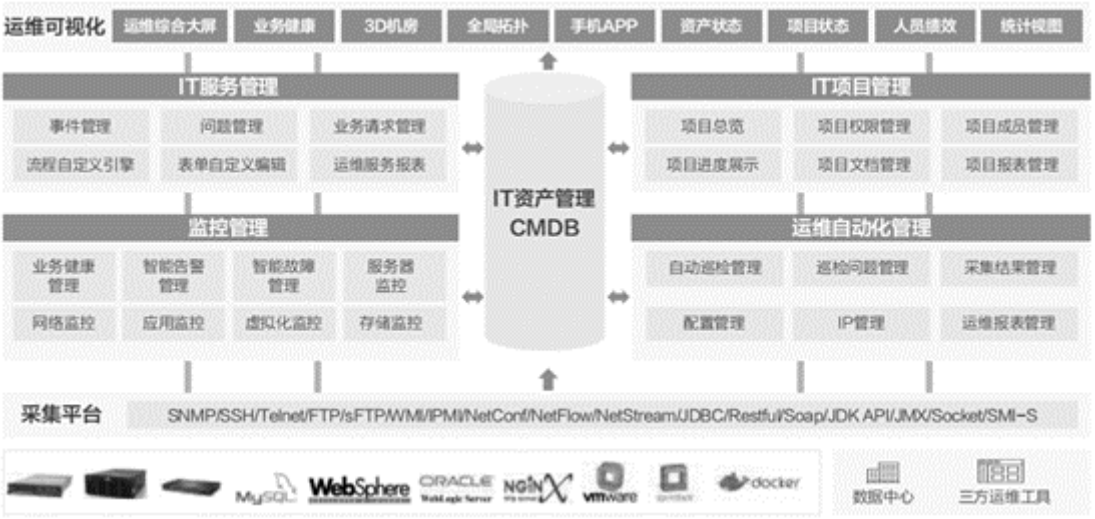
3 建设要点

数据是网络智能运维分析落地的基础，首先构建的就是网络基础大数据分析平台，运维数据进行采集、分 析、计算、存储，结合网络数字孪生模型和网络数据指标体系构建网络运维数据资源池，为上层运维分析业务应用提供数据支撑。路由器、交换机、防火墙、无线设备、终端设备等网络基础设施是网络运维分析的对象，是数据的主要来源，同时根据分析需要，也可以同步获取第三方网络管理运维系统的数据，让运维数据更加丰富完整，数据分析维度更为全面。数据资源池的数据分为基础数据和主题数据。系统通过对采集的设备数据、日志数据、流量数据进行提取、解析、转换和标准化处理后，形成资产信息、拓扑连接、配置数据、转发表项、流量数据、日志数据、认证数据、位置数据等运维基础数据集。运维基础数据按照业务分析主题进一步加工后形成各种主题库，如资产分析、拓扑分析、路径分析、应用分析、故障分析、容量分析等主题库。

结合医院的整体建设目标和思路，重新设计了新的运维思路，以 IT 资产管理（ CMDB ）为核心，通过多种手段对院内的各类 IT 资源进行统一的运行状态监控，并从业务视角对设备数据进行重构，实现业务层面的状态感知；同时将 医院的业务流程也逐步迁移到统一运维管理平台上，使得资源状

态的变更和数据的变化全程可管可控，打通人-资源-流程之 间的壁垒。

平台架构如下：



建设医院整体运维一张图初步预想：

4 预期带来的经济效益——挖掘信息部门的工作价值

✓ 提升工作效率

✧ 医院 IT 基础设施高度复杂和不断变化的情况下保持高质量的服务水平；

✧ 自动化流程工具极大提高医院 IT 服务部门工作效率，并降低人工导致的错误率；

✧ 提高患者满意度，提升门诊、护士站、住院等业务部门的服务体验；

✧ 极大缩短业务部门 IT 服务呼叫响应时间和解决时间；

✧ 以流程为导向，使 IT 管理的关注点从技术层面提升到流程和服务；

✓ 降低 IT 故障发生率

✧ 降低由 IT 基础架构中的错误引起的突发

事件和问题对业务的影响度；

✧ 降低故障发生率，极大提高 IT 运行的稳定性；

✧ 提高主动预防能力，在事件发生之前发现和解决可能导致事件产生的问题。

✓ 全面掌握 IT 信息

✧ 掌握医院 IT 基础架构中所有组件的最新的、准确的、全面的和详细的信息；

✧ 计量医院中所使用的所有 IT 资产和配置项的价值；

✧ 为支持流程提供帮助，提高解决效率；

✧ 保证 IT 基础信息的准确性和完整性；

✧ 核实有关 IT 基础架构的配置记录的正确性并纠正发现的错误。

✓ 控制 IT 风险

✧ 严格管控医院 IT 组件变更活动，有效降低变更可能导致的风险；

✧ 部门内部协调合作，维护变更过程的顺利进行。

✓ 降低运维成本

✧ 可向各业务部门提供自助式知识帮助提升了服务质量，降低了呼叫数量；

✧ 降低了新员工的培训成本；

✧ 使信息能够更快、更好的访问，从而提高了工作效率；

✧ 捕获有价值的知识，促进推动企业的创新。

✧ 利用自助服务向导，帮助用户自助分析和解决日常故障和问题，减轻 IT 部门的工作量；

✧ 业务部门和 IT 部门之间建立更加融洽的工作关系；

✧ 为 IT 管理提供了可量化的执行目标；

✧ 使 IT 部门的价值得到更好的体现，从而提高了员工的工作积极性；

✧ IT 服务提供方更为清楚地理解客户的需求，确保 IT 服务有效支撑业务流程。

✓ 工作绩效分析

✧ 减少人工统计工作，增强报表的准确性；

✧ 了解流程处理的效率，以及识别服务情况的趋势；

✧ 充分了解流程的宏观信息，掌控流程运作和组织情况；

✧ 发现流程的薄弱环节，为不断优化和改进提供指引；

✧ 为绩效考核提供基本依据。

福建医科大学附属第一医院作为国内首屈一指的医疗机构，在智慧医疗方面也一直是坚定的践行 者。构建统一运维管理平台是信息化战略中的一小步，通过新思路可以帮助医院更全面地掌控整个 IT 基础架构和业务的运行情况，并实现运维流程的电子化和标准化，为后续的智慧医疗的发展打下更坚实的基础。

5 未来展望

网络智能运维分析是网络运维与大数据和 AIOps 技术相结合的产物，是未来网络运维发 展的方向，AI 技术只是在局部场景中进行初步实践，可以通过实时获取网络运行状态指标和对异常指标进行检测，并实现部分故障场景的故障检测和自动恢复处置动作，实现分钟级故障检测、故障定位和故障自愈，有效提升运维人员的故障定位效率，保障网络的稳定 运行。

网络智能运维分析发展的终极目标是实现各运维场景的智能化闭环，实现运维数据资源池的标准化，故障场景标准化和故障处理自动化；网络智能运维分析系统和管理系统、SDN 控制器实现完全的一体化，并且和应用业务运维分析、ITSM 系统、运维自动化深度融合，AI 技术在故障模式识别、相似度关联分析、智能告警分析等领域实践落地，网络运维人员不再以发现和解决网络故障作为目标导向，有更多的时间和精力去探索网络如何更好地为业务提供服务支撑。

基于北斗网格码的涉密空间数据公众服务创新方案

李林[1]　任伏虎[2]　程承旗[2]
（1.北京大学智能空间（福州）创新实验室，福州 350007；
2.北京大学时空大数据协同创新中心，北京 100871）

摘　要:出于安全考虑，涉密测绘地理信息有严格的管理要求。但随着地理信息的广泛应用和信息技术的快速发展，数据安全（涉密）管理与数据公众服务的矛盾日渐凸显。在已有的火星坐标等做法之外，本文提出了一种兼顾数据安全与数据服务的创新方案，即：依托北斗网格码技术，将坐标变换为编码，并主要以“锚点+局部网格编码”的形式对外提供公众服务，避免了真实全球坐标的暴露，最大限度地扩展了服务范围，解决了现有跨地图系统位置难共享等问题。

关键词:北斗网格码；涉密空间数据；公众服务

1.问题的提出

1.1 涉密测绘地理信息管理要求

测绘地理信息是国家重要的基础性、战略性资源，广泛应用于经济建设、国防建设和社会发展，尤其是涉密测绘地理信息，直接关系国家主权、安全和利益[1]。各主要国家对于测绘地理信息成果均采取不同程度的保密方式。如：①美国实行军民分用，军用地图严格保密，民用测绘成果较为开放。采取地理信息要素分层方法，电子地图发布需经过政府审查。“9.11”事件后限制了敏感信息对公众开放程度，成立国家地理情报局；②英国：实施“敏感地点登记”，民用地图不标注登记在案的秘密军事基地。近年来政策有所放宽，但敏感场所不能标注名称，或使用含糊的名称；③俄罗斯：公开地图不得标示军用设施和敏感内容。对涉密成果提供要求严格。地图出版前要进行审查。

党中央、国务院历来高度重视涉密测绘地理信息安全管理工作。中央领导同志多次作出重要批示，要求落实总体国家安全观，进一步加强涉密信息安全保密监管工作，坚决防止测绘地理信息成果失泄密案件发生。目前，我国的涉密测绘地理信息管理已建立起相对健全的法律法规体系，即以《中华人民共和国测绘法》、《中华人民共和国保密法》、《中华人民共和国国家安全法》、《中华人民共和国数据安全法》为核心、相关配套规章制度为支撑的体系。

作为指导测绘地理信息工作的“基本法”，《测绘法》2017 年 7 月 1 日重新修订后实施，规定“地理信息生产、保管、利用单位应当对属于国家秘密的地理信息的获取、持有、提供、利用情况进行登记并长期保存，实行可追溯管理。从事测绘活动涉及获取、持有、提供、利用属于国家秘密的地理信息，应当遵守保密法律、行政法规和国家有关规定”；并专门增设“监督管理”一章，明确要求“建立地理信息安全管理制度和技术防控体系，加强对地理信息安全的监督管理，对属于国家秘密的地理信息的获取、持有、提供、利用实行可追溯管理，要求推广使用安全可信的地理信息技术和设备”。

2020 年 7 月自然资源部、国家保密局联合印发《测绘地理信息管理工作国家秘密范围的规定》（自然资发〔2020〕95 号），重新划分了测绘地理信息管理工作保密范畴、期限和密级，将涉密测绘地理信息成果明确为机密 12 项、秘密 14 项。简单

概括起来，常见的涉密范围包括：①“民用 1:1 万、1：5 千国家基本比例尺地形图；或多张连续的、覆盖范围超过 25 平方千米的大于 1：5 千的国家基本比例尺地形图及其数字化成果”；②“空间位置精度优于 50 米；影像地面分辨率优于 0.5 米的遥感影像（卫星遥感影像和航空遥感影像）”；③“平面精度优于 10 米或者地面分辨率优于 0.5 米、且连续覆盖范围超过 25 平方千米的正射影像”；④“平面精度优于（含）10 米或高程精度优于（含）15 米、且连续覆盖范围超过 25 平方千米的数字高程模型和数字表面模型成果”；⑤“平面精度优于（含）10 米或地物高度相对量测精度优于（含）5%、且连续覆盖范围超过 25 平方千米的三维模型、点云、倾斜影像、实景影像、导航电子地图等实测成果”；⑥“优于（含）20 米等高距的等高线，以及与其精度相当的高程注记点”；⑦“与军事、国家安全相关的国民经济重要设施精度优于（含）± 10 米的点位坐标及其名称属性”；⑧各类坐标转换参数（包含火星坐标处理算法及参数）。总体而言，强调的是真实坐标数据的保密和连续覆盖范围（25 平方千米）的控制。

1.2 数据安全（涉密）管理与数据公众服务的矛盾凸显

近年来，随着地理信息的广泛应用和信息技术的快速发展，涉密测绘地理信息安全管理面临新的挑战，原有的涉密管理技术体系正面临新的技术与应用形态的强烈冲击。如何解决涉密成果的保密要求与测绘地理信息面向社会的共享服务之间的矛盾，让公众能够安全便利地使用测绘地理信息成果？成为当前迫切需要解决的瓶颈性问题。

从需求角度看，物联网+大数据的浪潮正方兴未艾，社会的泛在感知体系逐渐形成，位置信息和位置服务受到越来越大的关注，对测绘地理信息成果的社会化需求前所未有的强烈。个人打车、外卖快递、物流配送乃至疫情防控等测绘地理信息新的大众化网络化应用呈持续高速增长态势。进一步剖析，现阶段地理信息数据生产、采集、利用均呈现多个新的特点，包括生产主体从专业化向大众化转变，服务内容从静态数据向网络动态数据转变，服务对象从以部门为主向以社会公众为主转变等。

总体而言，当前测绘地理信息载体种类和表现形式更加丰富，数字化成果广泛应用，传播途径更为多样，给涉密测绘地理信息的安全管理带来严峻挑战。特别是测绘地理信息数据在计算机网络上存储处理已成为常态，传统的以地图为信息载体、以单机封闭处理为主、严格限制信息传播范围和路径的涉密信息管理方式已经明显不能适应当前实践发展的需要，面向公众的网络地图和互联网地理信息服务等新形态已成为涉密测绘地理信息安全监管的最突出难题。

1.3 对地理信息保密处理技术的新需求

地理信息保密处理技术是现阶段保障涉密地理信息安全应用的关键技术，对维护国家地理信息安全、促进地理信息产业健康发展具有重要意义。[2] 目前最为广泛使用的是国家地形图保密处理技术（即通常俗称“火星坐标”），它是将各点的真实坐标施加不可逆的非线性平移变换，形成有偏差的非真实坐标，但又保证了任意目标图形的形状、大小、空间关系不发生变化。该技术自 2003 年投入使用以来，极大地促进了导航电子地图等产业的发展，对于增强测绘地理信息对全社会的支撑服务能力起到了积极的作用。但由于其是对不同厂商的地图分别进行加密变换，导致同一事物在不同地图系统里的位置不一致，带来很多使用中的不便。

因此，为满足日益增长的测绘地理信息公众化服务需要，适应“众包”社会化测绘、三维模型/实景数据/智能驾驶地图等高精度数据采集等新趋势的发展，需要探索既能满足数据安全保密又能最大限度实现数据共享服务的创新技术及解决方案。

2.创新技术：北斗网格码

2.1 体系内涵

北斗网格码技术是北斗网格编码与大数据组织利用技术体系的简称。[3] 它是一项中国自主原创、有望引领全球标准的时空大数据范畴基础性重大创新，包括两个层面内容：

2.1.1 北斗网格编码

北斗网格编码是一套新型全球空间位置框架

和编码方法，因被国家北斗系统列为新的空间位置输出标准而得名。其理论基础是北京大学程承旗教授团队承担国家 973 项目“全球空天信息剖分组织机理与应用方法”发展的新型地球空间剖分理论和大数据网格组织参考框架（GeoSOT 模型）。该模型将地表以上 52 万公里到近地心的地球全域空间剖分成最大为整个地球、最小 1.5 厘米的 32 级网格体元，每个网格均有唯一的二进制整形编码。

区别于传统的以经纬度二维指标定义平面位置，北斗网格编码创造性的以一维整形数定义三维空间，并辅以一套系统严密的计算规则，在应对数字孪生时代的空间信息处理具有独特优势。北斗网格编码具有多尺度立体性、超强计算性、良好包容交互性等基本特点。

2.1.2 基于网格编码的时空大数据组织利用

作为一种数据组织管理平台性技术，其将北斗网格编码自身面对的空间数据范畴扩展到具有时空属性的多源异构大数据范畴。首先，在数据组织层面，其将传统的面向对象的数据管理转化为面向空间的数据管理。无论对象数据的时态、结构等如何变化，利用空间网格的客观惟一性，以不变应万变；其次，在技术实现层面，利用空间网格剖分及时间剖分编码技术，建立统一时空数据组织框架，以时空数据为主索引可实现地球全域空间内万事万物数据的互联互通。

2.2 应用价值

作为时空大数据组织框架和大数据分析利用基础工具，北斗网格码技术体系能够支撑打造面向数字孪生世界的时空大数据底座，实现数字化新基建的万物数据互联互通，从而助力“数字中国、智慧社会”国家战略的推行。同时，随着北斗全球服务系统的开通，以北斗网格码为支撑的北斗数据增强服务正受到越来越多的重视，北斗网格码已成为推动北斗系统走向大数据的重要抓手，成为国家北斗应用战略的重要组成部分。

目前，基于北斗网格码的技术标准体系正逐步完善。国军标 GJB 8896-2017《地球表面空间网格与编码》、国家高分重大专项标准 GFB 30201-2018《高分卫星遥感信息剖分组织参考框架》、国家北斗标准 GB/T 39409-2020《北斗网格位置码》、国家标准 GB/T 40087-2021《地球空间网格编码规则》、国家标准 GB/T 40780-2021《基于 OID 的地理位置标识编码》已相继颁布；公安、住建、消防、邮政及智慧城市、空域管控等多个行业级标准已进入编制阶段。

3.基于北斗网格码的涉密空间数据公众服务方案

3.1 方案设计

3.1.1 设计思路

紧紧把握“敏感对象的真实坐标不暴露”这一核心目标，而不拘泥于“以精度定密、以范围定密、以比例尺定密”等传统做法，充分利用大数据、云服务等新一代信息技术，最大限度满足公众的空间数据共享服务需要。

核心的是建立一套基于网格的数据发布与共享服务体系。对于地理信息数据，将多尺度网格作为转换载体，通过锚点来串接真实全球坐标与局部相对坐标，形成统一服务平台支撑下的数据分布式发布与共享；对于基于地理形体的经济社会大数据，则直接由服务平台发布网格化数据（将原始数据按照一定规则处理后挂接在某一尺度的网格上）。

3.1.2 原理与优势

针对一个 30 千米*30 千米的地球表面局部区域，可近似视为平面处理。这基本能够覆盖城市级的主城区范围。选定某一锚点，对其全球坐标保密。再依托锚点建立高精度的相对位置平面直角坐标系，以支持局部精细化应用。这样锚点之外的各点真实全球坐标并不能推导出。同时，相对于火星坐标（相当于是移到错误位置），或低精度地图（信息不确定），信息正确性可以更好地得到保证，使得数据分析应用更具有科学性。

更为重要的是，实际应用中占据更大比重的是基于空间位置的社会经济大数据分析，其看重的是一定网格内的社会经济属性数据，而并不强调坐标数值的真实准确。通过网格这一数据载体，可一定程度跳出坐标数据保密的桎梏，基于网格而非坐标来发布数据服务；同时也可通过网格的粒度来实现

数据精度的控制。

3.2 实施步骤

3.2.1 将北斗网格码全球编码扩展到局部编码

根据实际应用指定一个锚点位置，根据北斗网格码编码规则可得到其全球网格编码（可与其真实全球坐标互逆换算）。再以此锚点为原点，根据其相对偏移量得到各位置的局部网格编码（单纯以该编码无法换算得出其真实全球坐标）。

3.2.2 城市公共锚点的设立并建立统一服务平台

在城市中设立 1 个或多个公共锚点（一级锚点）。该锚点的真实坐标不对外发布。建立城市统一的空间大数据服务平台。

3.2.3 面向用户的单次数据服务

根据用户的需求，城市空间大数据服务平台向用户定向提供数据或数据服务。数据内容由 1 个随机锚点（二级锚点）与若干相对位置编码（局部网格码）组成。该锚点由平台系统随机生成，锚点的全球真实坐标以及其与城市公共锚点的相对位置关系不对外发布。用户获得的相对位置编码可直接进行计算分析。

3.2.4 面向数据融合的服务

如不同用户所获得的数据或同一用户不同批次获得的数据需要融合，则可向服务平台发起申请。由平台将各批次的锚点及局部网格编码进行解算后，再作为一次数据服务按照前述 3.2.3 规则予以提供。

3.2.5 不同用户间的数据分享

不同用户间的数据分享，同样经由平台统一提供服务，用户可定向发送数据链接 URL 给到其他用户，而非脱离平台直接传递数据。

3.2.6 结合用户身份的数据传输加密

为提高安全可信水平，可进一步结合用户身份进行数据内容的加密传输，即：①随机生成单次密码；②将所要传输的数据内容按此密码进行加密；③将数据内容与密码分别传输；④用户按照收到密码进行解密，得到实际的数据内容。

3.2.7 网格化数据的公开发布

大量的数据需求并不在于空间数据本身，而在于基于统一空间尺度的对象内容数据，如该空间单元上的人口数量、建筑物面积以及车辆交通状况等。因此，空间大数据服务平台可按照一定的空间网格尺度公开发布这些网格化数据，为各方面的应用分析提供基础数据支撑。当然，这也就避免将原始数据直接部署到公共服务器可能产生的泄密风险。

综上，实际上是打造出一个城市级“后台+前端用户”的空间数据公众服务系统架构，即：（1）后台为可信受控环境，掌握全部真实数据；后台提供对外服务，根据锚点规则针对不同用户提供“随机锚点+局部网格编码”的数据内容；并根据加密规则针对不同用户每次生成随机密码，将数据进行加密传输。（2）用户须经身份认证；认证用户按需申请数据，接获数据包和密码后进行解密使用；用户间的数据共享通过平台来实现，避免用户间脱离受控环境的泄密风险。（3）在数据内容方面，既满足了数据安全保密的要求，又扩展了数据的公众服务能力，特别是实现了对“25 平方千米”等简单的面积限制的突破，能够支撑城市级乃至区域级的空间大数据分析利用。

4.该方案的创新价值

4.1 实现了真实坐标的保密

通过“锚点（真实全球坐标保密）+局部网格码（相对偏移量）”、“城市公共锚点+随机锚点”的方式对真实坐标数据进行了保密。不会出现一次性大范围坐标数据外泄，也避免了多次高重叠反推得到超保密面积的连续坐标，更不会泄露受保护敏感地物的真实坐标数据。

4.2 最大限度实现了数据共享

在锚点真实坐标隐藏的前提下，局部网格编码代表的相对位置关系准确，满足了计算分析等应用需要。特别是通过网格的方式可灵活进行多尺度的数据共享和数据发布，可打破“25 平方千米”的连续面积限制，便于城市级乃至更大空间尺度的大数据分析利用。

4.3 解决了跨地图系统的位置互认

由于火星坐标（国家保密插件）的应用，所有的地图/电子地图/导航设备都需要对真实坐标系统进行人为的加偏处理，且这个加偏并不是线性的。因此各地图系统间的位置往往不一致，也就出现居民按照互联网地图报警、公安按照警用地图出警但双方碰不上等情形。通过“网格+标志物”的方式可解决位置互认共享的问题。

4.4 提升了海量空间数据的计算服务能力

网格编码较之经纬度点坐标体系，在海量数据的计算效能上有数量级上的提升。北斗网格码技术体系和云服务平台架构的结合，能够支撑城市级时空大数据平台面向海量公众用户（及部门用户）的实时动态孪生化应用，能够更好地支撑实景三维中国、数字孪生城市等创新实践。

参考文献:

[1] 易树柏. 论地理信息安全在国家安全中的作用[J]. 理论界，2016,516(8):40-48.

[2] 贾宗仁. 从党的国家安全思想发展中深刻理解地理信息安全的发展与内涵[EB/OL]. 2021 年自然资源部测绘发展研究中心“学十九届五中全会精神，庆祝中国共产党成立 100 周年”主题征文优秀论文一等奖,2021.10.13.

[3] 李林，程承旗，任伏虎. 北斗网格码：数字孪生城市 CIM 时空网格框架[J]. 信息技术通信与政策, 2021,329(11):1-5.

基于大数据驱动的数据安全创新方案

陈新亮
（中国联通集团福建分公司，福建 福州 350001）

摘　要：随着国家监管机构对数据安全的高度重视，要求相关数据主责单位做好安全防护工作，特别是强调通过技术防护能力做好个人信息及重要数据保护，我单位立足基础电信行业合规要求，面向全省重要数据承载系统建立一体化的数据安全解决方案，达到集中管控、过程可视、自动扫描、智能稽核的应用效果，特别是针对数据导出风险创新性提出客户端导出操作在服务端自动留存数据副本，是对常规模式下仅日志留存机制的重大改进，改善风险事件数据追溯能力。

关键词：安全防护；重要数据；集中管控；创新性

一、方案背景

2019 年我单位针对工信部提出的基础电信企业数据安全合规性评估体系中涉及个人信息和重要数据保护相关的系统核心能力方面，组织专业团队对省内所有重要数据承载系统的数据安全防护能力现状进行了全面评估，基于我单位涉及重要数据的存量已达 10PB 量级规模，在对标能力缺漏及不足的基础上，要研究制定全省统一的、体系化的数据安全解决方案，力求方案在全面满足合规性管理要求的基础上，更加突出智能化、可视化、实用化、创新化，为中国联通集团福建分公司的整体数据安全筑牢坚强防线。

二、方案描述

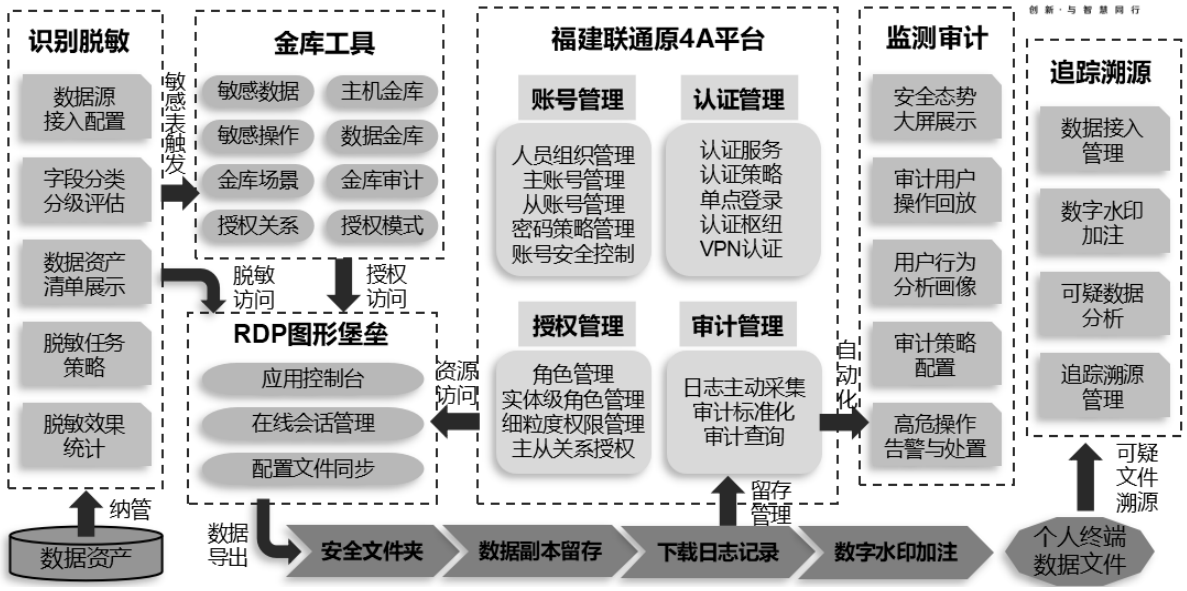
我单位在本方案实施前所面临的数据安全防护在能力上整体体现为功能欠缺较多、手工检查及处理环节较多、监控审计能力较弱等特点，仅部署有前端 4A 堡垒机平台进行基本的接入认证管控及基本操作日志告警及分析处理能力，在数据资产高效识别及分级分类、自动化审计、导出数据监控及追溯等方面还存在缺陷，无法实现企业整体的数据安全态势感知及满足端到端全闭环数据防护，主要的安全风险来自三个方面:

1. 重要数据保存较分散，数据资产管理手段空白,包括多类型数据库分散管理、字段信息不规范、资产清单台账缺失、敏感数据及重要数据缺乏分级分类打标的系统化管理等。

2. 基于数据全生命周期的管控手段不足,包括采集、传输、存储、使用及服务端数据导出等各个环节均或多或少存在管控手段覆盖不足。

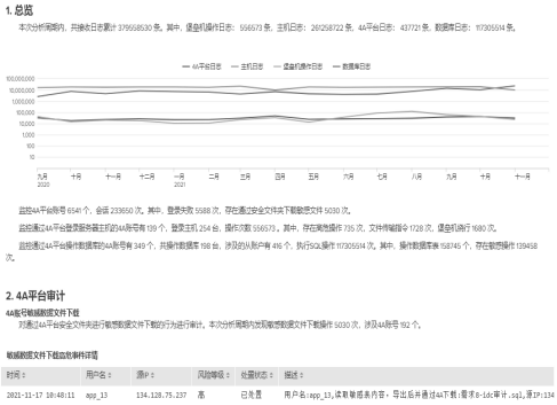
3. 数据安全治理手段不足， 特别是监控、稽核手段缺失，包括敏感数据异常操作、高频访问、调用 API 接口、实时化告警及事件触发等，数据导出虽然有流程审批，但实际导出操作在安全平台服务端没有留存导出数据的内容，无法追查样本数据，而依靠留存的操作日志无法进行高危样本数据的快速还原及追溯定位，对可能存在的高危数据导出操作震慑能力不足。

基于我单位的数据安全防护能力较薄弱现状及潜在安全风险，针对性进行方案的顶层设计,方案目标是建立覆盖全面、业务协同、上下贯通的数据安全技术防护体系，重点围绕资产分级分类、监测审计、识别脱敏、追踪溯源等场景，力求通过一个体系建立起企业整体的数据安全态势感知及管控能力，同时要求方案最大限度的利用好原有的 4A 堡垒机等硬件环境，主要通过叠加软件模块达成目标，保护原有资产投资。整体方案设计架构如下：

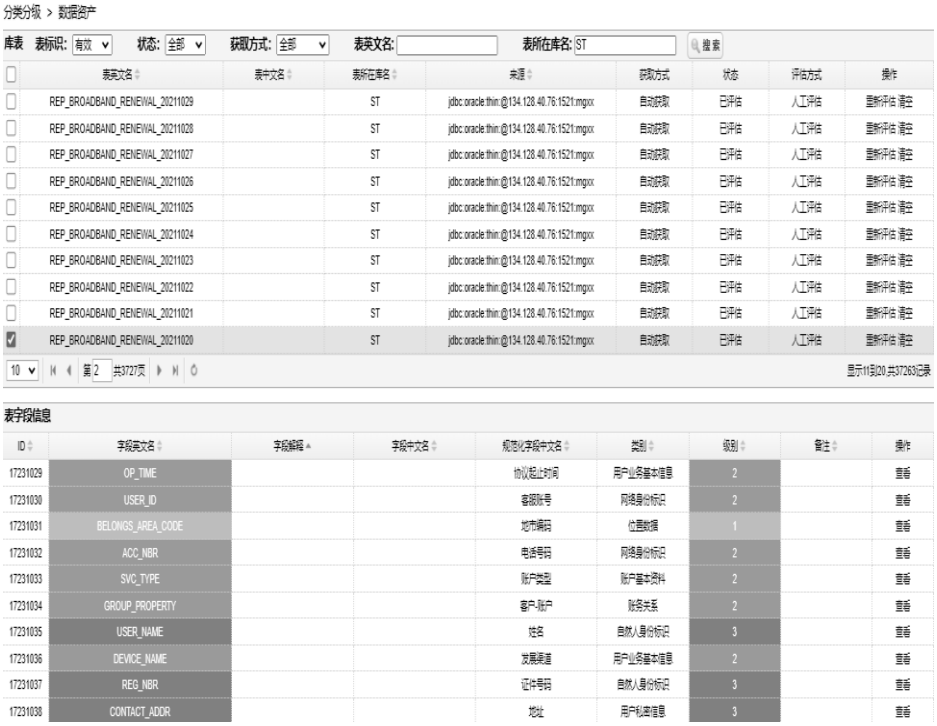


该方案设计思路主要包括以下四个核心模块：

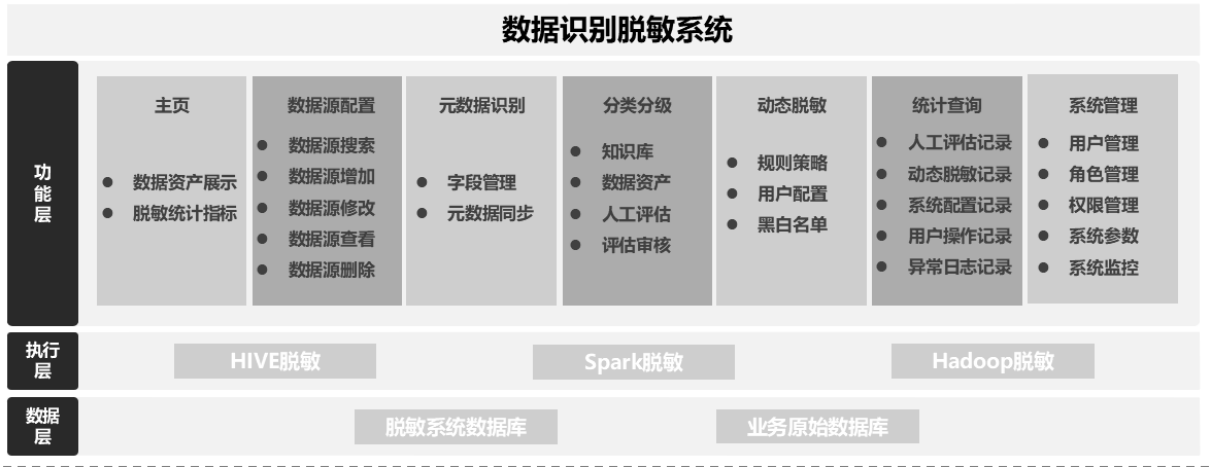
1. 监测审计:突出基于规则驱动的自动化审计能力，方案设计了基于关键词的策略配置中心，全可视化配置，可从主机操作数据的命令到数据库对象的操作脚本进行全面监测，对导出数据可按数据分级分类打标信息进行实时监测，审计报告可定期自动化生成，非常高效。对应我单位主机数据库规模及数据规模均较为庞大、数据访问场景及操作日志繁多的现实特点，强化基于大数据技术的处理能力，通过建立构建不同访问场景的基线监测模型，对采集到的各类操作记录进行自动化匹配及过滤，各类事件信息均通过可视化方式进行管理及展示，同时通过配置的触点实时发送监测信息到相应安全管理人员的手机或邮箱，通过实时化自动化提高管理效率。

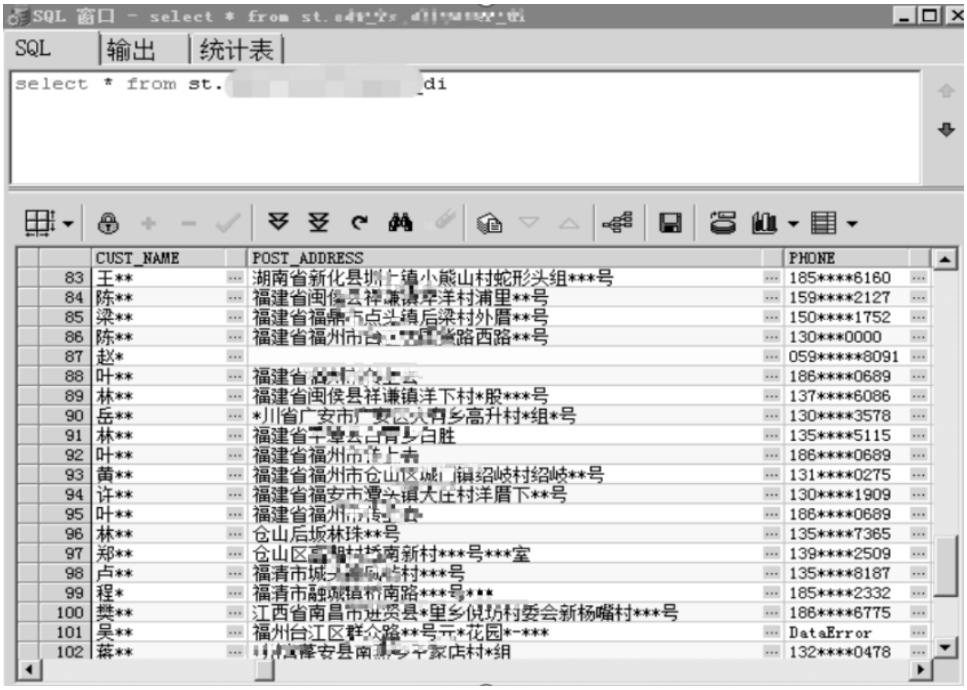


2. 数据资产分级分类：通过部署在服务端的智能化数据扫描工具，只须配置好相关重要数据承载系统的连接，同时调用内部的数据分级分类规则策略引擎，即可自动化扫描全量数据，能够发现识别敏感数据信息并形成资产台账，并按 4 级打标规则将分类资产进行入库。



3. 识别脱敏：采用基于内存的动态脱敏技术，根据全可视化的脱敏规则库配置、脱敏算法和脱敏软件组件，对数据源的元数据敏感度进行评估识别，针对重要敏感数据进行高精准度脱敏，确保各类数据处理场景中数据脱敏的有效性和合规性。通过脱敏算法策略动态匹配数据操作关键语句，实现对各大数据场景中的敏感数据脱敏展示。





4. 追踪溯源：针对可疑数据反向追溯数据的起源和经过的路径，获得数据流通的过程信息。数据主体搭载数据标识水印，有效追查数据泄露，震慑数据违规使用行为，为打击非法数据流通提供有效技术保障。通过创新性的将客户端的数据导出操作，在后台自动生成该操作对应的数据副本，以文件方式留存在服务端，对于企业生产运营现实需求中确实需要通过客户端进行数据库脱敏或非脱敏的数据导出行为，快速对疑似风险信息样本进行还原及操作追溯定位。服务端的数据留存，即对每个数据导出操作在服务端自动生成一份对应的数据文件，每个文件名对应一个唯一的 MD5 编码，确保每个数据导出操作都是唯一标识的，通过服务端对文件内容扫描快速定位出风险操作的数据文件内容，同时每个后台数据文件都记录匹配了具体的操作信息，数据在服务端留存采用文本压缩后占用空间小，几乎可长期在线提供安全管控使用。此种方式对比仅留存操作日志信息的方式，在风险数据内容的还原及定位上效率大大提升。


```
-rwxrwxr-x 1 nobody nobody      27311 Nov 18 08:42 1_cf33984a3c7c6020e624a370eb5a534c.csv
-rwxrwxr-x 1 nobody nobody    1468729 Nov 18 09:05 20211117_group_0点~24点_8fc09e5ca9ebf6c153b1d425101462ed.xlsx
-rwxrwxr-x 1 nobody nobody    7889325 Nov 18 09:09 202110账[illegible]正式稿（1118版）_4c05ac1caf57e0a38d25a382f3fe3071.xlsx
-rwxrwxr-x 1 nobody nobody    3628882 Nov 18 09:12 [illegible]其他1118_07b57e44ddab26f4aa15dab3146f5407.xlsx
-rwxrwxr-x 1 nobody nobody    1105230 Nov 18 09:12 [illegible]用1118_fe14eb987a3af0363520ff7018925323.xlsx
-rwxrwxr-x 1 nobody nobody    1248374 Nov 18 09:12 总部[illegible]结4G1118_c1e9834d1f0b2ad8e84f77f38b0b3c08.xlsx
-rwxrwxr-x 1 nobody nobody     122858 Nov 18 09:12 [illegible]表1118_78c2a0481d3433a51d6ed9a5d22f17b1.xlsx
-rwxrwxr-x 1 nobody nobody    5582336 Nov 18 09:13 202110账期总部现金补贴正式稿（1118版）_8862cf5682e9fa786d064b2393c106bc.xlsx
-rwxrwxr-x 1 nobody nobody       5769 Nov 18 09:18 20211118[illegible]_f8c1879434989dea87f16266ab19d98a.xlsx
-rwxrwxr-x 1 nobody nobody    5756119 Nov 18 09:21 11.18--202110账期[illegible]_180068de3854fbbf8efc767771400371.xlsx
-rwxrwxr-x 1 nobody nobody   12183376 Nov 18 09:24 11.18--2021[illegible]_015138d869e35ebc60aedb7521f5acf2.xlsx
-rwxrwxr-x 1 nobody nobody 1222071436 Nov 18 09:28 在网用户_快递预测_ac982d9f277cbd284f290a7213a03ffb.csv
-rwxrwxr-x 1 nobody nobody      74104 Nov 18 09:32 [illegible]明细_486cc5a2d6c66fe53e0dbd71a9d9472c.xlsx
-rwxrwxr-x 1 nobody nobody   51290798 Nov 18 09:32 中[illegible]站4_73c7e4768805515cd68f990302ca6f51.xlsx
-rwxrwxr-x 1 nobody nobody      94023 Nov 18 09:33 [illegible]汇总_465be2d589f64ac015892e672ee7e059.csv
-rwxrwxr-x 1 nobody nobody   17165739 Nov 18 09:34 迁[illegible]明细1118_3cb6749dbce271bd17309116ce8980e6.xlsx
-rwxrwxr-x 1 nobody nobody      79106 Nov 18 09:34 欠费延期明细_9e8935c1a24ae0b80fb300980951e6e6.csv
-rwxrwxr-x 1 nobody nobody        714 Nov 18 09:38 111_33d4c34758f6b0345a4ec93c33fc1868.sql
-rwxrwxr-x 1 nobody nobody   11424603 Nov 18 10:01 4_9d1d5930303d8dd9d502930a184219e7.xlsx
-rwxrwxr-x 1 nobody nobody       7179 Nov 18 10:03 15_4d29430a194424d6951a303df2a382b0.csv
-rwxrwxr-x 1 nobody nobody       7213 Nov 18 10:05 16_2fc5800972f7a0795b1b863d607266c4.csv
-rwxrwxr-x 1 nobody nobody    1287422 Nov 18 10:12 2_c632f3c18282df954b5eba0c69ebec0f.xlsx
-rwxrwxr-x 1 nobody nobody    2413444 Nov 18 10:12 3_f8e9c[illegible]c9a4bc339b234906c682c0e.xlsx
-rwxrwxr-x 1 nobody nobody     346630 Nov 18 10:14 20211118[illegible]_a131dd46273c98c94fb0ac307ed021d5.xlsx
-rwxrwxr-x 1 nobody nobody   17897421 Nov 18 10:36 标批语音产品用户1118_772324bc21a041779202d595dffa8df9.xlsx
-rwxrwxr-x 1 nobody nobody     467497 Nov 18 10:[illegible] 双[illegible]单1118_e08e3a943623e4fa6dd9553bc98396a1.xlsx
-rwxrwxr-x 1 nobody nobody     152882 Nov 18 10:47 [illegible]号清单_1bc21056bd532dc0b8cf80cf67dfdabc.csv
-rwxrwxr-x 1 nobody nobody       3537 Nov 18 10:48 1117_f8dee87bfa5a256d7c74259b1e59d4bc.csv
-rwxrwxr-x 1 nobody nobody   17519824 Nov 18 11:02 标批语音产品用户1118_def9855f1120dc716e67c9aa42eeb6fd.xlsx
-rwxrwxr-x 1 nobody nobody     491459 Nov 18 11:02 双[illegible]清单1118_3e932823de9c578dcee2aa594e8a4d63.xlsx
-rwxrwxr-x 1 nobody nobody    1378495 Nov 18 11:06 16日欠费回款数据_1fa427df9b7ce1526f213dac37823eae.xlsx
-rwxrwxr-x 1 nobody nobody       5411 Nov 18 11:12 通报1_7959662cf295a1d5d3e7fac1b46a5f84.xlsx
-rwxrwxr-x 1 nobody nobody       7329 Nov 18 11:13 通报2_b7bc3c52bd9e4147e4221b68fcdd1d80.xlsx
-rwxrwxr-x 1 nobody nobody       6008 Nov 18 11:13 通报3_6cdd8fc14f7a3d703e7955240eae6c4c.xlsx
-rwxrwxr-x 1 nobody nobody       6370 Nov 18 11:13 通报4_ed17cee964a1b8e359647e58709d16d0.xlsx
-rwxrwxr-x 1 nobody nobody       4793 Nov 18 11:14 通报6_cd674507e63ee6121a2916c30dfefd8a.xlsx
-rwxrwxr-x 1 nobody nobody   36172639 Nov 18 11:15 [illegible]帐1118_9308371980ae253e2d375f9f855f6e71.xlsx
-rwxrwxr-x 1 nobody nobody      46528 Nov 18 11:16 ST[illegible]_ALIPUSER_MI_891c33bf99df7c10203aff2506c4744f.tcl
[umap@localhost md5]$ pwd
/home/copy/md5
[umap@localhost md5]$
```

三、方案总结

本方案在 2020 年投入使用以来，在监测审计方面，共采集分析日志量达 21706608 条、审计模型策略配置达 268 条、高危风险核查处理 986 次；在资产扫描及动态脱敏方面，共纳管实体数据库各类对象 305409 项、分类分级打标字段共入库 4643653 项、配置脱敏策略模板 17 类； 在追踪溯源方面，经分平台数据文件水印加注 669 次，服务端的安全文件夹留存数据副本文件达 27013 项，可追溯所有的数据导出操作具体内容。方案设计在系统架构方面采用开放式平台架构、全配置化、可视化、模板化、产品化的设计思路保障了方案在适配不同行业的数据安全治理方面具有较高灵活性。

本方案部署后数据安全管理人员在安全管控及信息处理上的工作效率提升达 200%以上，自动化工作占日常工作的比重超过 70%，针对自动化运行输出，安全管理人员仅须进行数据抽查验证即可。我单位已连续多年保持重要数据泄露风险事件零发生， 连续两年在工信部的数据安全系统防护能力现场评估检查中保持满分，相关检查专家对本方案在满足合规性、创新性、实用性、可移植性方面给予了较高评价。

基于管理、技术、运营三位一体的数据安全防控建设解决方案

林　明
（福建中信网安信息科技有限公司 福建 福州 350014）

摘　要：构建一体化数据安全防控顶层框架，建立可持续化的数据安全运营体系，搭建闭环管控的数据安全技术体系，强化以数据安全技术为核心，以数据安全管理体系为指导，以数据安全运营体系为思路，构筑形成管理、技术、运营三位一体的数据安全防控框架，动态、持续地保障数据处理活动各个阶段安全有序开展。

关键词：数据安全管理；数据安全技术；数据安全运营；数据安全防控

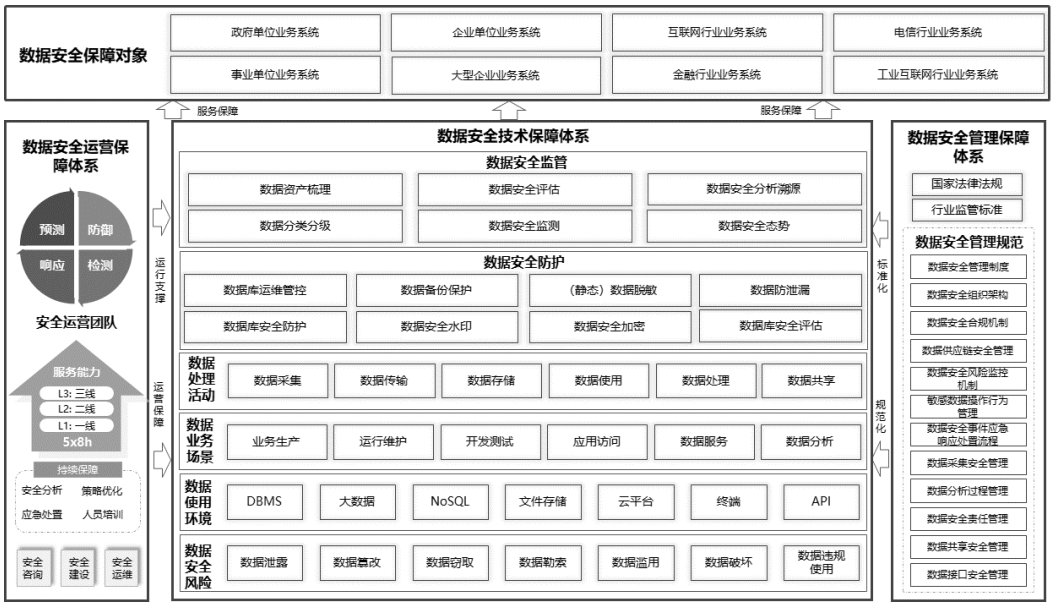
1 总体方案概述

坚持网络安全和信息化共同发展的理念，以保障政企事业单位、互联网、电信、金融、工业互联网等行业或大型企业业务数据安全为核心，进一步加强数据安全保障能力，建立有效支撑数据业务的安全体系框架，提高业务系统的数据安全监测、数据安全纵深防御、数据安全风险管控能力，全面保障业务系统和业务数据的安全。

对于保护业务数据的核心目标，一是要保障业务数据自身的安全，即为保护业务正常运行而必须保障业务数据的完整性、保密性和可用性，防止业务数据泄露或者被窃取和篡改。二是要保障社会重要业务数据的安全，在保障业务数据自身安全的同时，强化对重要业务数据的掌控能力，防止重要业务数据遭恶意的使用，对社会安全造成威胁。

2 方案整体框架

数据安全防控框架如下图所示：



通过在数据安全保障的过程中，充分考虑数据处理活动过程的所有环节，包括数据的采集、传输、存储、使用、处理、共享，通过数据安全保障体系实现数据安全的全部核心目标，即实现保护数据的完整性（防止数据篡改）、保密性（防止数据泄露、数据滥用）、可用性（防止数据破坏、数据勒索）。数据安全保障体系中通过建设使用数据安全监管及数据安全防护等相关技术手段对涉及数据的各种典型数据业务场景（如业务生产、运行维护、开发测试、应用访问、数据、数据分析）和数据使用环境（关系型数据库、非关系型数据、大数据环境、文件存储系统、云平台、终端设备、数据访问 API 等）中存在的外部数据安全威胁风险（如数据泄露、数据篡改、数据窃取、数据勒索、数据滥用、数据破坏、数据违规使用等）进行安全治理，最终全面有效的实现数据安全考查、评估、分析和防护。

3 数据安全管理保障体系设计

3.1 数据安全管理制度

根据业务、数据治理和信息化发展策略和目标，确定数据安全管控领域，建立安全策略、目标以及改进数据安全相关措施，形成数据安全管理办法，包含：目的、范围、岗位、责任、内外部协调机制及合规目标等。

3.2 数据安全组织架构

建立由决策层、管理层、执行层、监督层组成的数据安全管理组织架构。依据组织架构各层级管理职责，确定数据安全管理职能部门范围。

3.3 数据安全合规机制

调研外部数据安全及个人信息保护相关法律法规要求与标准规范，编制法律法规要求和标准规范清单，将满足和适用的条款、文件整理汇总形成数据安全合规资料库，同时依据数据安全合规资料库，梳理适用的评估标准，编制数据安全合规检查指标并全面覆盖检查。

3.4 数据供应链安全管理

调研数据供应链相关业务场景，梳理各业务场景中相关数据安全需求，汇总形成综合全面的数据供应链安全要求，建立数据供应链库，包括：完整的数据供应链授权信息、流转对账信息、场景使用信息等元数据信息。

3.5 数据安全风险监控机制

制定数据安全监控与审计相关规范及制度，明确监控和审计的策略、对象、内容、监控要求、异常流程处置，约定发生事件的处理流程等内容，制定对各类数据访问和操作的日志记录要求、安全监控要求和审计要求。

3.6 敏感数据操作行为管理

制定数据资源使用权限清单，包括大量敏感数据下载和查询的业务岗位、关键数据分析岗位、数据仓库访问权限设定，对人员访问数据资源进行访问控制和权限管理提供依据。

3.7 数据安全应急响应处置管理

对数据安全事件定义，评估业务数据和个人信息在数据应用过程中的泄露、滥用造成的影响，明确数据安全事件的处置流程和方法。制定完整的数据安全应急处置策略和要求，编制数据安全和个人信息应急处置策略和要求，编制定期组织开展应急培训和演练活动的流程。

3.8 数据采集安全管理

制定明确数据采集的目的、用途、方式、范围、采集源、采集渠道等内容，明确数据采集过程的数据保护要求，制定数据采集过程中的个人信息和重要数据的安全控制措施要求。

3.9 数据分析过程管理

规范数据分析全过程中的数据资源操作，覆盖构建数据仓库、建模、分析、挖掘、展现等方面，明确相关数据安全要求，确数据分析结果输出和使用的安全审核、合规评估和授权流程。

3.10 数据安全责任管理

建立数据使用者安全责任要求，明确说明违约责任、过失责任、侵权责任等，并说明在使用过程中采取的保护措施。建立数据管理员审计流程，明确对其操作合规的审计措施等。建立系统性梳理数据导出流程，包括识别未进行统一收口管理的数据导出操作的机制。

3.11 数据共享安全管理

明确数据共享原则及数据保护措施、数据共享涉及机构或部门的相关职责和权限、共享数据相关的使用者的数据保护责任。

3.12 数据接口安全管理

明确数据接口要求，包括从接口身份认证、防止重放、数据防篡改、防泄漏角度制定数据接口的安全限制和安全控制措施。明确对数据接口的安全限制措施，如身份鉴别、授权策略、访问控制机制、签名、时间戳、安全协议等。

4 数据安全技术保障体系设计

4.1 数据安全监管

4.1.1 数据资产梳理

对数据库进行定期主动的资产扫描及梳理，形成数据资产底账。基于建立的数据库数据资产底账，辅以人工刻画的管理域边界，对数据访问行为的分析，动态侦测数据资产的变化情况，发现数据库中存在的僵尸资产、复用资产、复活资产和失踪资产。对数据库账号及权限进行监测分析，对数据安全责任边界进行界定，动态刻画数据安全边界，明确数据的使用者、责任者和所有者。

4.1.2 数据安全监测

数据安全监测包括：业务数据安全监测、API 接口数据安全监测、运维数据安全监测、开发数据安全监测。

（1）业务数据安全监测

对业务数据库账号越权访问行为、业务数据跨库异常访问行为等行为进行监测；

（2）API 接口数据安全监测

对 API 业务接口进行资产发现梳理、变更安全监测、数据行为审计监测；

（3）运维数据安全监测

对非法终端异常访问数据库行为、业务器直连数据库异常行为、运维数据库账号越权访问行为、运维数据风险操作行为及业务敏感信息违规触碰行为进行监测；

（4）开发数据安全监测

对开发测试阶段的数据异常抽取行为、敏感数据异常访问行为及数据库账号越权访问行为进行监测。

4.1.3 数据安全评估

在数据处理活动中，通过对可能发生对于数据保密性、数据完整性、数据可用性造成危害的威胁行为进行识别，掌握数据资产的脆弱性面临威胁的可能性，并判断对数据资产的影响程度及数据安全风险的大小。同时，对现有的数据资产安全措施进行识别并对其有效性进行分析确认。

4.1.4 数据安全态势

内部员工的恶意破坏、违规操作和越权访问，往往会带来数据的大量外泄和严重损坏，甚至导致数据库系统崩溃，这些操作往往不具备攻击特征，很难被普通的信息安全防护系统识别。通过提供业务数据资产风险态势、数据安全态势、数据库运维态势，实现全域数据安全态势感知，动态识别现有数据安全状态。

4.1.5 数据安全分析溯源

以数据安全事件为入口，以数据风险模型为基准，对安全事件进行回溯和调查，可视化绘制出完整的事件生命周期，包括的源、目标、途径、范围等相关信息。

4.2 数据库安全防护

构建数据库安全防护技术措施，提供数据库级别访问控制、入侵防御，有效的保护后台数据库不暴露在复杂的网络环境中，构建数据库的安全防护状态。

4.3 数据库运维管控

构建数据库运维管控技术措施，提供数据库运维人员进行访问、操作数据库进行管控，规范数据库运维行为，防止内部违规操作、误操作，避免数据泄露、丢失。

4.4 （静态）数据脱敏

构建（静态）数据脱敏技术措施，提供对数据库中的敏感数据进行识别，统计出敏感数据和管理敏感数据，实施敏感数据脱敏处理，同时保证数据的有效性和可用性，使脱敏后的数据能够安全的应用于测试、开发、分析，和第三方使用环境中。

4.5 数据安全水印

构建数据水印技术措施，提供数据分发的安全防护，通过数据水印技术加以保护，用增加伪行、伪列队原始敏感数据中嵌入不易察觉且难以去除的标记，在不破坏原有数据内容和对象的可用性前提下，实现保护数据安全的目的。

4.6 数据库安全评估

构建数据库安全评估技术措施，提供数据库的漏洞扫描、配置核查、弱口令检测，及时发现数据库存在的安全漏洞，通过及时加固，保障数据库自身的安全。

4.7 数据安全加密

构建数据加密技术措施，提供主动的数据安全防御机制，防止明文存储引起的数据泄密、突破边界防护的外部黑客攻击和内部高权限用户的数据窃取行为，同时防止绕开合法应用系统直接访问数据库的外部攻击和窃取。

4.8 数据防泄露

构建数据防泄露技术措施，提供主动的文件存储防泄露和网络防泄露的数据泄露防护机制。对留存在电脑终端、服务器、文件共享和数据库上的敏感数据信息进行识别，对文件共享、邮件、Web、应用程序等传输的数据进行监控，在数据进行操作之前对其进行管控。检查网络传输的流量，解析各种会话流量协议，识别传输的敏感数据并进行强制数据防控控制策略，阻断非授权的敏感信息传输并及时进行告警。

4.9 数据备份保护

构建数据备份技术措施，提供数据库数据的备份机制，有效避免出现数据因各种原因遭到删除而无法恢复的情况。

5 数据安全运营保障体系设计

5.1 人员驻点运维

提供人员现场驻点，开展数据安全日常的安全检查、安全事件处理等工作。

5.1.1 数据资产发现管理

对数据资产进行持续维护，同时发现敏感数据库和敏感数据的位置和分布，统计重要数据，并对数据资产的归属部门、责任人、使用方等信息进行备案登记，形成数据资产列表。其次，定期对数据库资产进行持续的安全评估，从而识别数据库资产的脆弱点并作为后续数据安全优化的基础。

5.1.2 数据分类分级管理

基于数据分类分级策略模板，借助敏感数据发现工具及数据自动分类分级工具，对新发现的数据资产进行分类分级识别与打标，不断完善现有的数据分类分级清单，并进行人工核实确认。

5.1.3 数据账号、应用、终端资产管理

收集分析数据资产访问情况，并从中提出数据相关的访问行为特征及对象，从而识别出应用数据库账号、数据库运维账号、数据访问应用、数据访问终端等，不断完善数据库账号、应用、终端资产清单，并由人工进行核实完善。

5.1.4 数据安全合规管控

梳理评估各个级别的数据资产在整个数据安全生命周期中的可能存在安全风险，并借助平台评估行为的安全报告中整改措施，由人工对针对不同的数据风险设定数据安全防护节点的防护策略，从而对数据安全进行合规管控。

5.1.5 数据安全风险持续监测

对数据风险进行持续监测，持续不断的监控数据处理活动中存在的数据流转情况，并针对产生的数据安全事件进行告警。

5.1.6 数据安全事件响应处置

对数据安全事件提供响应处置，同时落实针对数据安全事件应急响应措及处置措施，并能够后续针对数据安全事件进行溯源取证。

5.1.7 数据安全规范、制度、办法优化

针对建设的数据安全管理规范，当业务流程发生变更时进行相关数据安全规范、制度、办法等优化，持续保障数据安全管理规划的完整性、合理性及准确性。

5.2 二线安全支撑

二线支撑专家接受来自一线驻点人员的支持

请求，负责对安全事件进行分析，找出事件产生的原因并确定解决方案，从发生的事件中找出事件的发展趋势或潜在可能发生的问题，提供预防性措施建议，提高数据安全可靠性。

5.3 三线安全技术专家

通过组建三线安全技术专家团队，提供数据安全顶层技术支撑，解决一线安全驻点工程师、二线安全支撑团队无法解决的问题，协助解决驻点运维过程的各种数据问题，提高信息部门的数据安全事件响应与处理能力。

6 方案总结

通过建设智能高效的数据安全防控机制，协助掌控业务核心数据资产，健全全域数据资产底账，明确各业务数据的价值和责任。健全数据安全一体化追踪手段，进一步升华核心数据全生命周期安全管控防护手段，从粗放被动审计升级为精细化数据全流程主动防御，保障核心数据安全可控，全面提升数据安全监管防护能力，降低核心数据泄露风险并满足数据安全法和个人信息保护法的合法合规要求。

参考文献:

[1]中关村网络安全与信息化产业联盟数据安全治理专业委员会.数据安全治理白皮书 4.0 [R],2022.

[2]全国人民代表大会.中华人民共和国数据安全法[Z], 2021.

[3]全国人民代表大会.中华人民共和国个人信息保护法[Z], 2021.

[4]中国信息通信研究院.大数据白皮书2021[R], 2021.

数据安全管控平台

冯晓敦
（中国电信股份有限公司福建分公司，福建　福州 350001）

摘　要：随着《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》的颁布实施，对数据安全共享和管控的要求日益提升，福建电信建设数据安全管控平台，实现了对数据流动安全监测、数据库防火墙、数据防泄漏等数据安全管控原子能力的统一管理，提供了数据应用安全防护、数据风险识别与处置、数据脱敏、数据操作管控、安全审计等技术工具手段，同时采用大数据 AI 技术，实现数据的自动化分类分级与重要数据标识，形成动态更新的企业数据安全台账，把识别结果作为数据能力输出给各种数据安全管控工具，实现场景化融通应用，提升数据安全管控效率，减少人工运维工作量，实现了对数据安全的可视、可控、可管，提升了数据安全管控水平。

关键词：数据安全；分类分级；大数据 AI；安全能力；融通应用

1 概述

1.1 目标客户群体

随着《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》的颁布实施，对数据安全共享和管控的要求在法律法规层面上做了规定。《数据安全法》规定“各地区、各部门应当按照数据分类分级保护制度，确定本地区、本部门以及相关行业、领域的重要数据具体目录，对列入目录的数据进行重点保护”；《个人信息保护法》在个人信息处理者的义务中明确要求要对个人信息实行分类管理。

数据安全分类分级管控已经成为各行业各部门都需要实施的重要工作要求。数据安全管控平台的目标客户群体，包含政企事业单位、互联网、电信、金融、医疗等各行各业，特别是存放用户个人敏感信息或者重要数据信息的行业，更是存在建设完善数据安全管控手段的迫切需求。

本方案采用通用的系统架构，按照模块化、松耦合的原则设计，数据安全管控各能力功能模块都相对独立，接口开放兼容性强，可以方便实现和其他安全工具的对接。各行业各部门都可以根据各自所处的数据安全管控发展不同阶段、不同需求，选用相关功能模块实现统一对接和管理，迅速构建符合本企业需求的数据安全管控各平台，实现对数据安全的可视、可控、可管。

1.2 方案拟解决的问题

本方案充分应用大数据 AI 技术，完成数据的分级分类自动化精准标识，节约了大量人工识别标志的工作量，提升运维人员支撑工作的效率，特别是存放海量数据的行业部门，数据分类分级的智能化自动化处理是必不可少的手段。

本方案提供了丰富可选的数据安全管控工具能力。结合数据安全监测模型，建设多个数据安全能力，包含流动数据安全监测、数据库防火墙、数据防泄漏等数据安全管控原子能力等，并可以实现对这些能力工具平台的统一编排、统一调度、统一管理。数据安全管控平台提供开放的接口，安全工具能力之间各自独立，客户可以根据单位部门的实际需求，选择其中的 1 项或者多项能力，进行灵活部署。

本方案可以把数据分类分级识别结果作为数据能力输出给各种数据安全能力工具进行场景化应用，作为这些工具实现自动化安全管理的数据底座，并根据这些工具的应用情况来持续修正完善安全数据

质量。通过融通应用，可以自动化实现具备本行业本单位特点的个性化的数据安全控制要求，减少人工配置操作，持续提升数据安全管控工作的自动化和智能化水平，减少人工运维工作量。消除数据安全信息孤岛，构建融通一体的数据安全防护圈。

2 方案说明

2.1 数据安全管控平台统一门户

作为企业数据安全的统一运营门户，包含了资产管理、策略中心、事件监测、风险分析以及大屏展示等功能。

2.1.1 资产管理

资产管理功能包含了数据资产发现和数据资产管理,其中数据资产包含了数据库(含大数据组件)、文件系统、接口、用户账号等内容。

数据资产发现主要是指管控平台能够通过流量探针、端口扫描等方式发现未知数据资产，可以通过数据资产发现中的内容判断是否进行认领，长期未认领的资产将转化为未备案风险，提示用户进行下一步的处置。

数据资产管理主要是指管控平台可以通过手动新增、批量导入、接口同步、对接第三方管理系统（例如资产管理系统）、对接大数据湖等方式获取全量数据资源信息，从而集中备案并管理企业内的所有数据资源信息。

2.1.2 策略中心

策略中心的功能主要是实现平台上所有策略的统一管理，策略包含了资产管理分类分级、数据安全监管、关联风险分析等策略集中统一管理。

资产管理分类分级，通过平台的统一策略配置能够实现数据资产的自动发现、敏感数据的扫描识别、分类分级规则的配置等一系列关于资产和分类分级的策略。

数据安全监管，数据安全管控平台集成了流动数据安全、数据库防火墙、数据库审计子系统、数据防泄漏以及数据脱敏、数据存储加密等安全组件的策略配置，实现了统一平台管理多组件的模式，极大的方便了用户对企业内数据安全组件的管理。

风险分析，通过数据安全管控平台汇聚的各种安全日志信息，通过对日志归并计算、行为基线、AI 模型、数据算法等的配置，实现一个或者多个日志的联动，从而发现深层次的数据安全问题。

2.1.3 事件监测

事件监测模块主要将底层各安全组件的日志、告警以及第三方系统的日志、告警汇聚，并提供展示和查询。通过该模块既可以在一个界面内总览、查询、处置所有数据安全组件的原始告警事件，又可以分别查看单一安全防护组件的事件和日志。该模块还提供多种对接方式，支持个安全组件、业务

系统、管理系统或者大数据湖通过 Kafka、syslog、接口同步等多种方式将告警日志、操作日志、用户登录日志等数据安全相关日志汇聚于数据安全管控平台，通过事件监测模块，能够看到各个安全组件上报的告警事件和日志。

2.1.4 风险分析

风险分析模块主要基于事件监测模块汇聚的大量安全事件和日志信息,突破单一组件告警的瓶颈，通过对海量的日志信息进行数据清洗和分析，通过行为基线、AI 模型、大数据统计等方式，将离散的日志信息进行关联计算，从而发现潜在的数据安全风险，并提示客户。

通过日志关联分析能够发现单一系统或者单一安全防护能力无法发现的问题，特别是通过多个日志的关联能够很好的还原数据安全问题链条，更好的定位源头或者责任人，能够从根源上解决企业内部潜在的数据安全风险。

2.1.5 大屏展示

大屏模块主要用于总览整个平台的资产管理、策略中心、事件监测、风险分析等各个模块的统计信息，包含了三个大屏，分别是综合态势、资产态势和风险态势。

其中综合态势主要是展示平台的整体情况，包含了纳管数据资产的基本情况、分级分类的大盘、安全组件策略的总体情况以及事件和风险的总体情况等。

资产态势主要集中于资产和分类分级相关数据的展示上,区分了多个维度来对上述信息进行展示，包含了数据资产在各业务系统的分布情况，敏感数据的分布以及流向等，让客户能够更好的了解企业内的资产和敏感数据分布以及分布流向情况，能够通过该大屏更好的了解自己企业的敏感数据情况。

风险态势主要展示了总体的安全事件情况、风险情况以及由此产生的工单以及工单处理情况。

2.2 数据安全能力中心

数据安全能力中心作为数据安全监测防护的重要组成部分,包含了各类数据安全防护和监测能力，同时各能力通过数据安全管控平台统一管理。

2.2.1 流动数据安全监测

流动数据安全监控系统对流量镜像后进行旁路解析并对流量进行双向还原，实现数据流动安全风险全面有效治理，实现流动环境中的应用、接口、IP、账号等自动梳理,流动的敏感数据的动态监测，更好的帮助客户对流动数据的监测和风险的防范。

流量探针支持对开放 API 流量、应用系统流量、数据库流量、文件访问流量等数据源通过旁路、代理、agent 等方式进行流量采集解析、双向流量还原，支持对异常风险事件进行一键封堵，支持接收和同步各能力层下发的数据采集监控策略。

智能化引擎模块，提供多种能力引擎包括应用

账号解析引擎、风险事件监测引擎、文件还原引擎、UEBA 建模分析引擎、敏感数据分析引擎、异常事件封堵引擎、信息溯源引擎等。

2.2.2 数据脱敏

可对人员、权限、客户端、主机、时间等不同维度配置脱敏策略，针对待脱敏数据可以进行替换、屏蔽等方式进行脱敏处理。确保不同团队根据不同权限访问数据信息，保护敏感信息，不被违规查询泄露，提升数据安全性。

2.2.3 数据库审计

基于数据库通讯协议准确分析和 SQL 解析技术，实现了对数据库操作、访问用户及外部应用用户的审计，可以用于安全合规、用户行为分析、运维监控、风控审计、事件追溯等与数据库安全相关的管理活动。

2.2.4 数据库防火墙

通过 SQL 协议解析技术实现对 SQL 语句以及表列级的防护，能够实时监测和阻断 SQL 注入攻击、数据库漏洞攻击以及脱库撞库等外部黑客攻击行为；可以规范内部用户的访问行为，阻断内部高危操作如访问系统表，不带条件的更新和删除。可以弥补数据库审计产品只能事后追溯的弊端，满足等级保护要求，为用户构建数据安全主动防御体系，保障用户敏感数据安全。

2.2.5 数据防泄漏

通过流量牵引技术，对受控区域内的外发流量进行深度解析、内容恢复和敏感度扫描，及时发现受控区域内通过网络泄漏数据、传播数 据的行为，并进行拦截、告警、审计等措施，能够根据网络环境和监控需求，进行灵活多变的部署。

3 方案创新性、先进性和成效

3.1 应用大数据 AI 技术，实现数据分类分级的自动化精准标识

应用大数据 AI 技术，在数据资产中精准区分敏感数据与非敏感数据，通过内置 AI 机器学习算法规则和内置行业法规标准，基于深度学习+条件随机场的命名实体识别模型，可以准确、高效的识别，并自动对各业务系统存储的数据进行分类分级。产品支持丰富的敏感数据识别技术：包括正则表达式、关键字典、机器学习、NLP、文档指纹等先进 AI 技术创建识别规则、实体识别模型等。形成动态更新的企业数据安全台账。

3.2 采用 DPI 等多项技术实现数据安全动态信息的全量采集解析

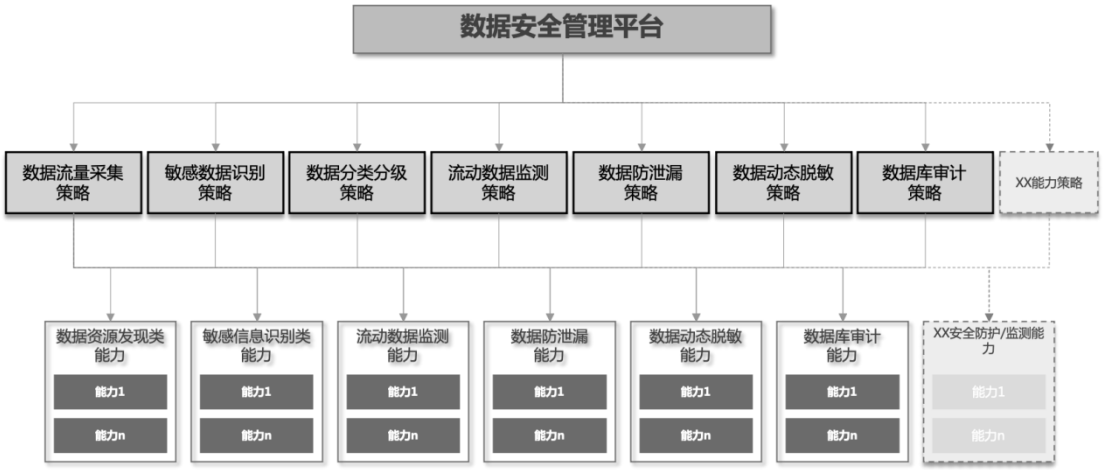
采用DPI等多项技术实现数据安全动态信息的全量采集解析，支持全流量审计，实现对关键字、数据来源等的自定义，通过内容深度匹配流量中的敏感信息，并对敏感信息快速定位，实现对敏感信息访问行为的有效监测。

3.3 基于 UEBA 构建数据安全动态检测模型，实现数据安全风险的智能化预测分析和管控

结合大数据、AI 等技术，构建数据安全动态检测模型，多维度、多层次关联分析识别风险，实现数据安全风险的智能化预测分析和管控。采用大数据和机器学习技术，对多维度的信息和数据进行整合、关联、智能分析和预测，辅助安全人员做出最精准的判断和调查。基于 UEBA（用户行为分析），采用 ARIMA 差分整合移动评价自回归模型和 NLP 自然语言处理算法对每个用户或实体的历史访问行为进行学习、训练和建模，形成各种访问行为基线和预测值，监测程序再根据这个预测值对访问行为做出预判，如果实际值和预测值偏差太大，则预判为异常，这为部分风险场景自动化预警提供了精确有效的判断依据，适用场景如：访问流量异常、单 IP 访问频次异常等。

3.4 开放的数据安全管控平台，实现数据安全能力、数据安全运营管理工作的统一管理和控制

数据安全管控平台具备能力开放接口，实现数据安全能力、数据安全运营管理工作的统一管理和控制。平台研发策略转换器，当安全能力纳入数据安全管控平台集中管控时，平台通过策略转换器自动将策略转换为目标能力支持的运行规则，具体原理如下所示。



通过策略转换器使得数据安全管控平台能够对管理的数据安全能力进行统一的策略编排、统一的能力调度，同时数据运营管理工作事项和流程，也可以通过平台进行固化、流转，从而实现数据安全管控、运营一体化。

3.5 数据能力和安全能力工具平台的融通应用

把数据分类分级结果作为数据能力，输出给各安全能力工具平台，作为这些工具平台实现自动化安全管控的数据底座，实现对大数据 MBO 等业务系统的数据安全管控的场景化应用，并根据这些工具平台的应用情况来持续修正完善安全数据质量。通过融通应用，减少人工配置、干预，提升自动化水平，让数据能力和安全能力充分发挥作用，提升数据安全管控效率，减少人工运维工作量。消除数据安全信息孤岛，构建融通一体的数据安全防护圈。

3.6 成效

随着 2021 年两法的颁布实施，数据安全管控相关工具手段的研发，在国内各行业都是刚刚起步、方兴未艾，此方案无论在省内还是电信集团内部，都具有较高的创新性和先进性，并取得良好的应用成效。中国电信对全集团 2021 年 9 月数据安全日评分中，此方案的实施助力福建电信公司取得满分，位列集团前列。本方案获评 2021 年福建电信公司“省级优秀案例”。本方案获评 2021 年福建电信公司转型创新二等奖。本方案，经福建电信公司评选通过，上报参与工信部 2022 年网络安全技术应用试点示范工作。此方案目前已经在福建、河南、吉林三省的电信公司全省推广和应用，取得良好的应用成效。

医院网络安全分析与规划方案

林传捷
（福建医科大学附属第一医院 信息中心，福建 福州 350001）

摘　要：探讨医院网络系统安全现状与规划。方法：分析医院实际环境中的网络安全问题。结论：针对现有系统的缺陷并提出全面性、系统性的解决方案

关键词：网络安全；等级保护；应用系统；管理体系

1 网络安全挑战

网络安全工作是医院信息化建设的重要内容和关键环节。从 2019 年 5 月，网络安全等级保护标准 2.0 正式发布，到《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》等法律法规密集出台，对医院网络安全提出了更高的要求。我们根据网络安全策略对医院网络系统进行全面的分析，从通信网络、区域边界、计算环境等方面对网络安全进行了全面评估，主要面临的网络安全挑战如下：

1.1 通信网络安全挑战

1. 针对网络架构设计不合理而影响业务通信或传输问题，需要通过优化网络设计、改造网络安全域来完成。

2. 针对线路或设备的单点故障问题，需要采取冗余设计来确保系统的可用性。

3. 针对利用通用安全协议、算法、软件等缺陷获取信息或破坏通信完整性和保密性，需要通过数据加密技术、数据校验技术来保障。

4. 针对通过伪造信息进行应用系统数据的窃取风险，需要加强网络边界完整性检查，加强对网络设备进行防护、对访问网络的用户身份进行鉴别，加强数据保密性来解决。

1.2 区域边界安全挑战

区域边界包括安全计算环境边界，以及安全计算环境与安全通信网络之间实现连接并实施安全策略的相关部件，区域边界安全即各网络安全域边界和网络关键节点可能存在的安全风险。需要把可能的安全风险控制在相对独立的区域内。具体如下：

1. 针对内部人员未授权违规连接外部网络，或者外部人员未经许可随意接入内部网络而引发的安全风险，以及因使用无线网络传输的移动终端而带来的安全接入风险等问题，需要通过违规外联、安全准入控制以及无线安全控制措施来解决。

2. 针对跨安全域访问网络的行为，需要通过基于应用协议和应用内容的细粒度安全访问控制措施来解决，以实现网络访问行为可控可管。

3. 针对通过分布式拒绝服务攻击恶意地消耗网络、操作系统和应用系统资源，导致拒绝服务或服务停止的安全风险，需要通过抗DDoS攻击防护、服务器主机资源优化、入侵检测与防范、网络结构调整与优化等手段来解决。

4. 针对利用网络协议、操作系统或应用系统存在的漏洞进行恶意攻击（如碎片重组，协议端口重定位等），尤其是新型攻击行为，需通过网络入侵检测和防范等技术措施来解决。

5. 针对通过恶意代码传播对主机、应用系统和个人隐私带来的安全威胁，需要通过恶意代码防护技术手段解决。

6. 针对邮件收发时遭受恶意代码攻击的安全风险，需要通过垃圾邮件防护等技术手段解决。

7. 针对违规越权操作、违规访问网络等用户行为，需要采取安全审计手段来实现安全事件的有效追溯和用户行为的审计分析。

1.3 计算环境安全挑战

计算环境安全涉及业务应用系统及重要数据处理、存储的安全问题。具体安全挑战如下：

1. 针对用户帐号权限设置不合理、帐号暴力破解等等安全风险,需要通过帐号管理、身份鉴别、密码保护、访问控制等技术手段解决。

2. 针对在网页浏览、文档传递、介质拷贝或文件下载、邮件收发时而遭受恶意代码攻击的安全风险，需通过恶意代码防范、入侵防范等技术手段解决。

3. 针对操作用户对系统错误配置或更改而引起的安全风险，需通过安全配置核查、终端安全管控等技术手段解决。

4. 针对设备系统自身安全漏洞而引起被攻击利用的安全风险，需要通过漏洞扫描技术、安全加固服务等手段解决。

5. 针对攻击者越权访问文件、数据或其他资源，需要通过访问控制、身份鉴别、安全审计等技术来解决。

6. 针对利用各种工具获取应用系统身份鉴别数据，进行分析获得鉴别内容，从而未授权访问、使用应用软件、文件和数据的安全风险，需要采用两种或两种以上鉴别方式来，可通过应用系统开发或第三方辅助系统来保证对应用系统登录鉴别安全；

7. 针对应用系统缺陷、接口设计等导致被恶意攻击利用、数据丢失或运行中断而影响服务连续性的安全风险，需要通过对产品采购、自行软件开发、外包软件和测试验收进行流程管理，同时保证应用软件具备自我容错能力；

8. 针对由于应用系统存储数据而引发的数据损毁、丢失等数据安全问题，需通过本地数据备份和异地容灾备份等手段来解决；

9. 针对个人信息泄露的安全威胁，采取必要的安全保护手段；

1.4 安全管理中心挑战

1. 针对系统管理员、审计管理员、安全管理员的违规操作行为，需要采取角色权限控制、身份鉴别、安全审计等技术手段对其操作行为进行限定，并对其相关操作进行审计记录。

2. 针对众多网络设备、安全设备、通信线路等基础设施环境不能有效、统一监测、分析，以及集中安全策略分发、恶意代码特征库、漏洞补丁升级等安全管理问题，需要通过集中安全管控和集中监测审计机制来解决。

3. 针对应用系统过度使用服务器内存、CPU等系统资源的行为，需要对应用软件进行实时的监控管理，同时对系统资源进行管控来解决。

4. 针对设备违规操作或多通路运维带来的安全风险，需要对指定管理区域及安全管控通路。

1.5 安全管理体系挑战

1、安全管理制度涉及安全方针、总体安全策略、安全管理制度体系、评审与修订管理等方面。

2、安全管理机构涉及安全部门设置、人员岗位设置、人员安全管理等方面。

3、安全运维管理涉及环境管理、资产管理、系统安全运行维护管理、配置与变更管理、安全事件处置及应急响应管理等方面。

2 医院整体规划

2.1.设计目标、依据、原则

2.1.1 设计目标

在统一的安全保护策略下要具有抵御大规模、较强恶意攻击的能力,抵抗较为严重的自然灾害的能力,以及防范计算机病毒和恶意代码危害的能力；具有检测、发现、报警及记录入侵行为的能力；具有对安全事件进行响应处置,并能够追踪安全责任的能力；遭到损害后，具有能够较快恢复正常运行状态的能力； 对于服务保障性要求高的网络，应该能够快速恢复正常运行状态；具有对网络资源、用户、安全机制等进行集中控管的能力。

2.1.2 设计依据

本方案设计主要参考国家、行业信息安全指导政策、标准方法与最佳实践，包括但不局限于：

- 卫办综函[2011]1126 号（《卫生部办公厅关于全面开展卫生行业信息安全等级保护的通知》）
- 闽卫信息函[2012]63 号（福建省卫生厅福建省公安厅关于印发《福建省医院信息系统安全等级保护工作实施方案》的通知）

- 公信安〔2009〕1429 号文件 《关于开展信息安全等级保护安全建设整改工作的指导意见》
- GB 17859-1999《计算机信息系统安全保护等级划分准则》
- GB/T 20984-2007 《信息安全风险评估规范》
- GB/T 22239-2019 《网络安全等级保护基本要求》
- GB/T 22240-2008 《信息系统安全保护等级定级指南》

2.1.3 设计原则

信息系统安全配套建设应参照国家等级保护、ISO17799 和 IATF 等标准，综合考虑可实施性、可管理性、可扩展性、综合完备性、系统均衡性等方面因素，在信息安全设计过程中应遵循下列原则：

- 整体性原则

进行安全规划设计时应充分考虑各种安全配套措施的整体一致性。

- 符合性原则

信息安全体系建设符合有关国家技术标准，以及行业的技术标准和规范。

- 均衡性原则

安全体系设计要正确处理需求、风险与代价的关系，做到安全性与可用性相融，寻找安全风险与实际需求之间的一个均衡点。

- 有效性与实用性原则

信息安全系统不能影响业务系统正常运行和合法用户的操作。在进行网络安全策略设计时，要综合考虑实际安全等级需求与项目经费承受能力的因素。

- 等级性原则

对业务系统的不同单元进行信息保密程度分级，对用户操作权限分级，对网络安全程度分级（安全子网和安全区域），对系统结构分级（应用层、网络层、链路层等），针对不同级别的安全对象，提供全面、可选的安全算法和安全体制，以满足各不同层次的实际需求。

- 统筹规划、分步实施原则

信息安全防护策略的部署既要考虑满足当前网络系统及信息安全的基本需求，也要统筹考虑后续系统的建设及网络应用的复杂程度的变化，做到可适应性的扩充和调整。

- 动态化原则

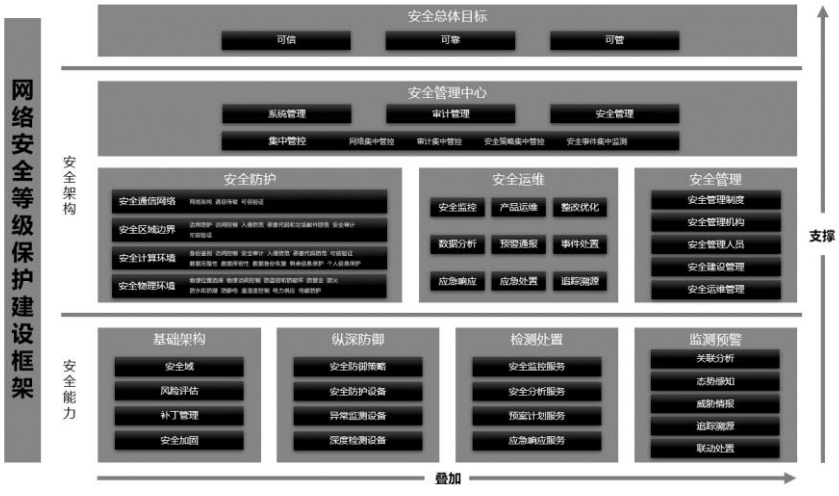
随环境、条件、时间的变化，安全防护策略不可能一步到位，信息安全系统应能适应变化，采取更先进的检测和防御措施，增强安全冗余设备，提高安全系统的可用性。

2.2 总体医院安全规划

2.2.1 方案设计框架

信息系统安全等级保护体系框架在国家政策、法律法规要求的指引的前提下，以安全基础设施为依托，与信息系统的业务流程、应用架构和数据资源紧密结合，以技术、管理为要素进行框架设计。

网络安全等级保护建设框架如下图所示：



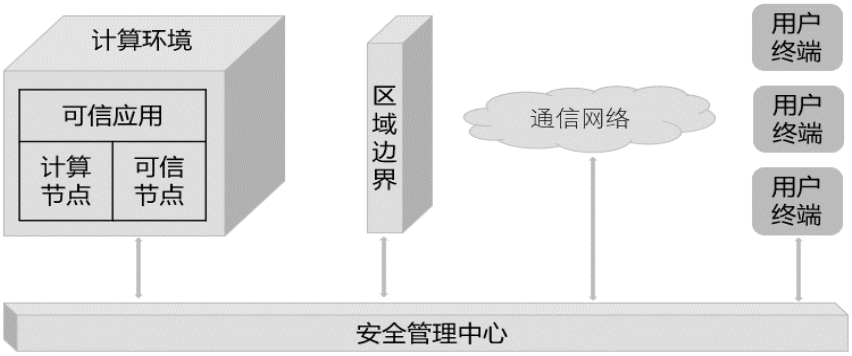
在安全体系方案设计时，将根据国家信息安全等级保护相关要求，通过分析系统的实际安全需求，结合其业务信息的实际特性，并依据及参照相关政策标准，设计安全保障体系方案，综合提升信息系统的安全保障能力和防护水平，确保信息系统的安全稳定运行。具体设计将遵循以下思路开展。

1、构建安全基因的设计思路

合规要求与业务风险分析相结合，信息安全风险分析是识别信息系统面临安全威胁和系统脆弱性的方法，通过风险分析方法可以全面掌握信息系统面临安全风险的全貌，并根据安全风险等级确定信息安全建设的重点，在完成基于资产风险分析的基础上，对信息系统现状进行实际调研，掌握系统防护现状与等级保护基线要求间的实际差距，结合信息安全风险评估的方法，对信息系统进行全面的资产、脆弱性、威胁和业务风险等方面系统化的评估分析，发现基于业务的安全风险问题。将差距分析结果与风险评估结果进行充分结合与提炼，综合形成能够符合等级保护建设要求并充分保障业务安全的建设需求。

2、纵深防御的设计思路

信息系统安全体系建设的思路是根据分区分域防护的原则，按照层次化的纵深防御的思想，建设信息系统安全等级保护深度防御体系。



按照信息系统业务处理过程将系统划分成安全计算环境、安全区域边界和安全通信网络三部分，以计算节点为基础对这三部分实施保护，构成由安全管理中心支撑下的计算环境安全、区域边界安全、通信网络安全所组成的“一个中心，三重防护”结构。

3、持续保障与改进的设计思路

结合安全防护措施、安全管理制度、安全运维服务，实现对信息系统的多层保护，及持续保障与改进的目标。

4、监测预警积极防御的设计思路

基于持续性的风险监测预预警、追踪溯源、联动处置，结合采集威胁情报、安全态势感知，实现网络安全积极防御目标。

2.2.2 安全通信网络设计

依据等级保护要求第三级中网络和通信安全相关安全控制项，结合通信网络安全审计、通信网络数据传输完整性/保密性保护、可信连接验证等安全设计要求，安全通信网络防护建设主要通过网络架构设计、安全区域划分、流量均衡控制、通信网络安全传输、通信网络安全接入，及通信网络安全审计等机制实现。

2.2.3 安全区域边界设计

依据等级保护要求第三级中安全区域边界相关控制项，结合安全区域边界对于区域边界访问控制、区域边界包过滤、区域边界安全审计、区域边界完整性保护及可信验证等安全设计要求，安全区域边界防护建设主要通过基于地址、协议、服务端口的访问控制策略；非法外联/违规接入网络、抗DDoS攻击、恶意代码防护、入侵防御、APT攻击检测防护、无线安全管理以及安全审计管理等安全机制来实现区域边界的综合安全防护。

2.2.3 安全计算环境设计

依据等级保护要求第三级中安全计算环境相关控制项，结合安全计算环境对于用户身份鉴别、自主访问控制、标记和强制访问控制、系统安全审计、用户数据完整性保护、用户数据保密性保护、客体安全重用、可信验证、配置可信检查、入侵检测和恶意代码防范等技术设计要求，安全计算环境防护建设主要通过身份鉴别、安全访问控制、安全审计、入侵防范、恶意代码防护、主机可信验证、数据完整性保护、数据保密性保护、个人信息保护、数据备份恢复以及系统和应用自身安全控制等多种安全机制实现。

2.2.5 安全管理中心设计

依据等级保护要求第三级中网络和通信安全相关安全控制项，结合安全管理中心对系统管理、安全管理和审计管理的设计要求，安全管理中心建设主要通过运维审计、网络管理系统、综合安全管理平台等机制实现。

2.2.4 安全管理体系设计

信息安全管理体系是组织在整体或特定范围内建立的信息安全方针和目标，以及完成这些目标所用的方法和体系。以等保安全管理基本要求为基础，结合 ISO270001 的体系的 PDCA 过程和 ISO27002 的 14 个控制域规范，同时兼顾监管部门的相关安全规范，整合企业自身的 IT 服务管理体系和安全技术防护体系，通过体系规范化、管理流程化、测量指标化、操作工具化的手段来确保体系设计的落地。

3 结论

通过医院网络安全分析规划，根据相应的保护等级的要求进行规划设计，通过通信网络、区域边界、计算环境、安全管理中心及安全管理体系的设计，提升医院信息系统网络安全保护水平。

5G 网络安全空间测绘

谢 辉
（中电福富信息科技有限公司，福建 福州 350000）

摘　要：5G 网络空间测绘是构建网络安全体系的底层支撑。其中核心要点之一就是要做好网络空间资产的测绘工作，通过扫描探测、流量监听、主机代理、特征匹配等方式，动态发现、汇集数据，并进行关联分析与展现，以快速感知安全风险，把握安全态势，从而辅助用户进行指挥决策，支撑预测、保护、检测、响应等安全体系的能力。

关键词：测绘；安全；防护；网络空间；威胁检测

背景

为了应对网络安全风险带来的挑战，国家层面不断加强对网络安全工作的政策指导，法律法规和标准层面不断完善安全防护体系和指南，产业发展层面不断明确差异化的安全需求。多方的安全诉求也为 5G 安全提出了多层次的防护要求，必须将 5G 安全作为 5G 生态的组成部分予以充分考虑。

从国家政策要求来看，党和国家对网络安全工作的重视程度不断提升。党的十九届五中全会提出，要“统筹发展和安全，建设更高水平的平安中国”，要“坚持总体国家安全观”、“统筹传统安全和非传统安全，把安全发展贯穿国家发展各领域和全过程”。5G 作为新基建之首，传统安全和非传统安全融合，安全要求自然首当其冲。

从法律法规和标准来看，《网络安全法》、《关键信息基础设施保护条例》、等保 2.0 系列标准以保护国家关键信息基础设施安全为重点，提出了网络安全实战化、体系化、常态化新理念，注重全方位主动防御、安全可信、动态感知和全面审计，强化安全集中管控，要求持续增强动态防御、主动防御、纵深防御、精准防护、整体防控、联防联控六大能力，并对使用新技术的信息系统提出了安全扩展要求。5G 作为数字经济的重要基石，需要在安全保障更为可靠有效、安全响应更为快速准确等方面提供更为有力的支撑。

1 解决方案

产品将网络空间、地理空间和社会空间进行相互映射，将虚拟、动态的网络空间测绘成一份动态、实时、可靠、有效的网络空间地图，为决策者提供有价值的战略情报信息，降低决策的不确定性。打造实现一个精准、实时、智能的网络空间安全态势感知体系。构建“全资产管理、探测、深度感知、情报预警与可视呈现”的网络空间安全资产测绘平台，实现对终端主机、设备、流量、协议、控制系统等的全方位实时探测与感知，基于轻量化、无感知的探测技术与独有关联分析算法对海量的资产数据进行融合关联分析，结合风险告警和趋势预警技术，形成经安全指数综合评估的实时多维度安全态势与威胁情报，便于及时处置资产安全风险和威胁。

提升网络安全监管能力：通过全网资产安全监测及数据融合分析，实现对网络空间资产安全的在线实时态势感知，提升用户网络资产的安全监管能力；

提高网络安全防护水平：通过平台的安全态势感知的风险预警机制及时向用户进行网络空间资产风险与威胁通报，督促系统或资产管理部分加强安全防护，提高整体网络安全防护水平；

提高合规性执行监督能力：通过网络安全等级保护合规性，实现对网络空间资产安全合规性执行

情况进行追踪，提升监管部门对网络安全合规性监督能力。

目前产品通过如下图所示的产品架构，通过数据采集、数据分析、数据存储、数据处理、数据应用、数据呈现等方式，实现了将网络空间资源属性以及网络资源间的关联关系进行建模和表达，实现全网网络空间全要素全息数字化映射和可视化视图呈现，以反映网络空间资源状态变化、网络行为。

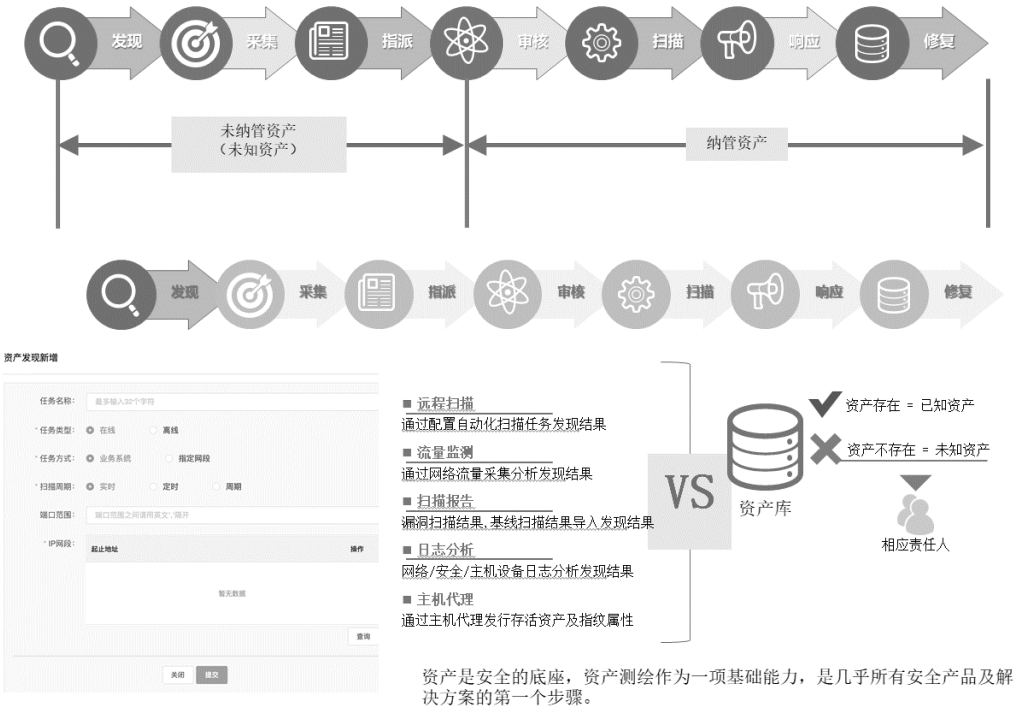
结合目前的网络技术和数字化发展愿景，网络空间测绘的应用场景主要包括网络安全应用领域、数字化管理应用领域以及测绘定义网络领域。随着技术进步、时间发展和应用的不断深化，这三个方向是层层递进的，目前产品的功能建设规划也是根据这三个方向来进行规划建设的，主要解决方案如下：

a) 5G 资产测绘

资产的发现能力，互联网地址的快速广泛应用以及 5G、物联网等技术发展使得越来越多的资产需要暴露在互联网空间，更大的网络空间暴露面带来了更多的网络空间安全威胁。如何摸清家底，持续探测网络空间资产,如何时刻洞察网络空间资产，主动掌控资产动态，发现未知或未备案的资产，并能够动态的采集数据源的 IP、端口、服务、类型、版本等等属性，形成资产指纹库，这些是资产探测的核心能力。通过多种方式发现网络空间资产。

b) 5G 资产生命周期管理

通过远程扫描、网络流量、日志分析、扫描报告、主机代理实现资产的初步探测及发现，并对未知资产进行全生命周期管理，对发现结果与现有资产库比对，主动发现未纳管资产并生成纳管工单，完善资产库及对应指纹库，否则很多成为三无七边资产，存在一定安全隐患。



c) 5G 资产绘制

绘资产数据的关联性，资产发现获取的资产信息，繁杂且不规则，通过对资产信息的分类分级、标签标识、关联分析、安全风险等等，使其成为更有价值的资产数据，这个过程中形成的能力就是对资产的关联绘制，资产测绘的主要应用场景，是将资产数据与威胁风险关联叠加后，提升安全应急响应的时效，从而能够预防并快速处置网络空间安全隐患。

d) 5G 资产脆弱性分析

管理和监控重要 5G 资产存在的脆弱性信息。各种重要的主机、终端和网络设备上存在的安全脆弱性是影响信息安全的重要潜在风险，实现对重要主机系统和网络设备安全脆弱性信息的收集和管理，对收集的信息进行分析，形成安全事件，驱动工单系统处理安全事件。

e) 互联网暴露面核查

从攻击暴露面的角度来看，随着传统的资产范畴以主机资产、网络设备为主，信息资产，再到如今 App、API、微信公众号等新的数字资产形态，资产的范畴不断外延，暴露在潜在攻击者面前的选择越来越多。而这些数字资产在传统台账中恰恰是完全缺失的，如果仅依靠安全管理员使用传统方式手动发现、完善、补充，必然会有遗漏。因此安全厂商的资产测绘能力要能涵盖这些新的数字化资产。具备互联网暴露面核查能力，发现暴露公网的 IP、端口、进程及异常行为。

f) 5G 资产入侵检测

处理各类入侵事件及具有高度威胁的事件，支持识别并处置的入侵威胁事件包括：病毒木马、网页后门、反弹 shell、异常账号、日志删除、异常登录、异常进程、系统命令校验等。对接国内外主流查杀引擎，可检测出恶意进程及软件，并提供隔离、信任等功能。

g) 5G 资产病毒防护

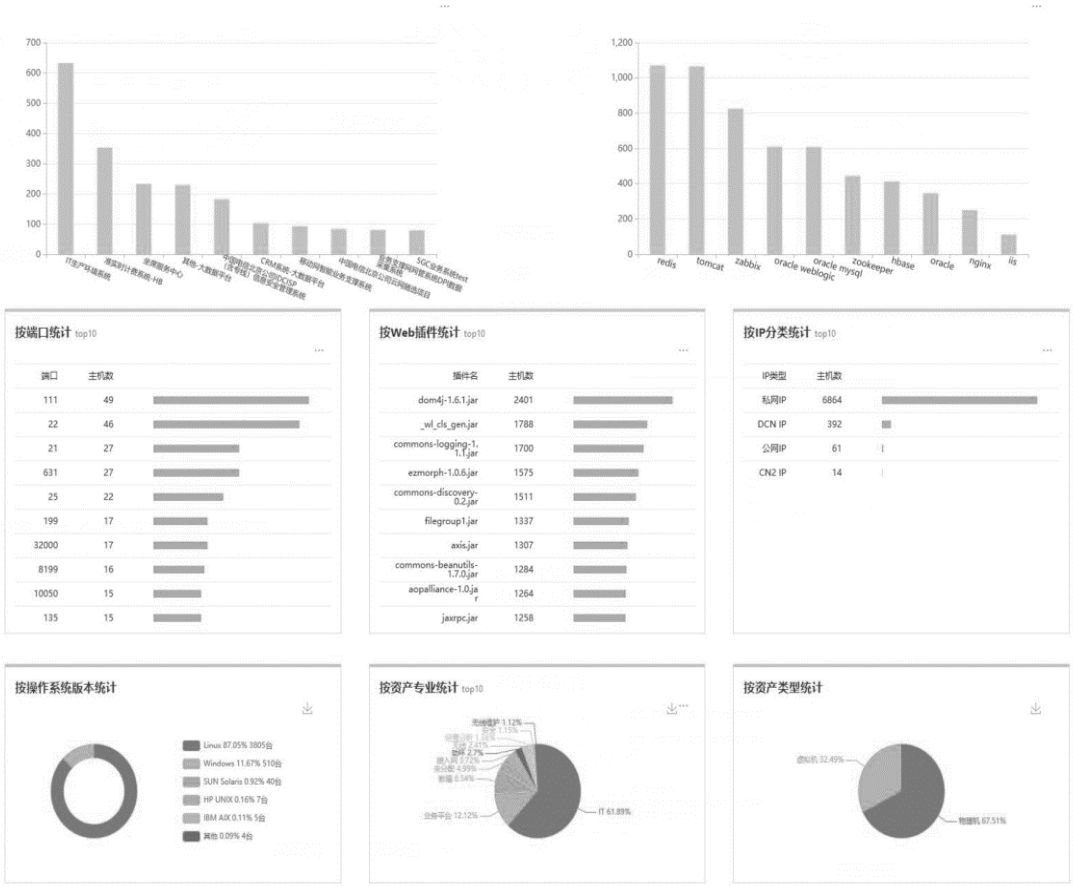
采用防病毒引擎和病毒特征库相结合，是去检测和发现病毒的程序。而病毒库是已经发现的病毒的标本。用病毒库中的标本去对照机器中的所有程序或文件，看是不是符合这些标本，是则是病毒，否就不一定是病毒。

h) 风险关联

重点梳理资产暴露面，全方位掌握资产的暴露面，摸清所有的关联关系及安全风险，从攻击视角出发，将资产暴露面的梳理涵盖到更广的范围，掌握自身资产状态，定位存在漏洞的风险资产，全面排查安全隐患资产。

i) 多维度溯源分析

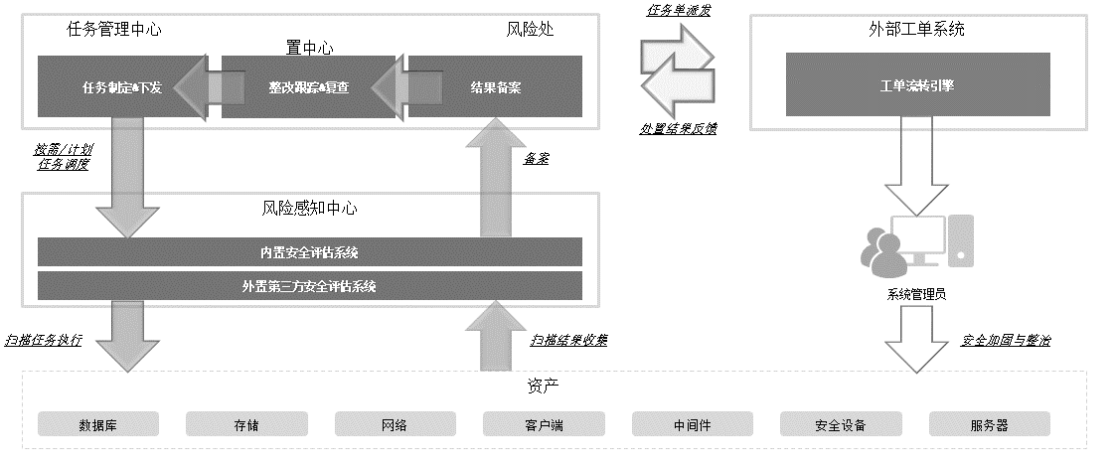
利用数据分析引擎，对数据源、采集数据、识别结果等内容进行全面分析，包括统计分析、价值分析、关联分析等，实现企业数据资产的全面测绘，形成企业数据资产地图、多维统计分析视图、资产分析报告、资产清单等。



j) 5G 资产风险处置

■ 编排处置

联动风险处置模块，针对发现的资产安全风险进行闭环全流程管控，其中响应方式包括工单下发、邮件以及短信等方式；支持一键封堵关停，从源头切断威胁。如一键封堵处置管理模块主要根据告警处置内容对路由器（或主机代理）模块下发封堵指令，该指令为 IP（IP+端口）地址。



■ 闭环处置

以在线或者离线任务的方式，创建实时/定时

/周期性任务对已纳管资产进行扫描评估，并能够自动识别资产所的 IP 类型，避免有限扫描资源的浪费；评估结果发现资产脆弱性后，能够进行漏洞的的备案、下发、整改、复扫等一系列闭环处置流程；个性化的业务系统 KPI 考核模式能够针对每个自然周期内业务系统的新增资产、漏洞、整改率等多维度进行关联分析，并对相关业务系统责任人进行考核。

2 目标客户

网络空间资产的基础地位决定，无论是监测全网态势还是行业监管或企业自身的安全防护，网络空间资产测绘工具都有着巨大的发展空间。所以在应用层面，工业互联网、物联网、5G 通信、IPV6 等信息环境，包括大量智能终端、网络设备和软件应用等计算对象的爆发，以及现实世界与网络世界的融合，会衍生出大量的应用场景，空间测绘一定会在其中扮演极为重要的角色。可以在这些行业应用层面推广试用。从技术层面，空间测绘将走向主动探测、被动分析、探针代理等多种手段结合，并结合自动化，甚至是人工智能等技术，以适应未来庞大、复杂的应用场景需求。所以标准化产品，可快速部署试用，可复用性大，扩展性强，适合于规模推广。

覆盖 5G 场景：

5G 资产管理：对 5G 资产进行全方位采集分析，发现探测资产指纹库，识别资产风险、脆弱性核查、互联网暴露面分析、入侵检测及病毒防护等等进行关联绘制，预防并快速处置网络空间安全隐患。

5G 工业互联网：工业互联网资源测绘和安全分析能力，绘制工业互联网重要资源地图和安全图谱，以实现对工业互联网基础设施的有效探测、空间感知和安全分析、风险预警和损害评估。构建工业互联网资源测绘和安全分析平台，发现工业互联网设备、系统、平台等重要资源，通过多点联动的协同发现、融合分析等，形成跨地区、跨行业、跨领域工业互联网资源。

5G 智慧城市：新一代智慧城市是数字化的重要场景。需要通过集中建设和运营“城市级”网络安全基础设施提升城市网络安全水平。网络空间测绘地图作为网络安全基础设施中的重要组成部分，能够为城市网络安全管理者清晰展现数字世界的虚拟资源，帮助用户高效管控网络空间，维护城市的网络安全。

3 创新点

平台采用后续规划采用高可用的大数据安全自适应架构，轻松处理数 T 级别数据，秒级处理威胁检测。快速部署，平台最重要的是可靠稳定，可快速在客户环境中落地部署、展现效果。所以标准化产品，可快速部署试用，可复用性大，扩展性强，适合于规模推广。

全面的安全信息收集，通过多种标准协议或定制的收集工具全面收集安全设备、网络设备、主机系统等各类设备产生的日志数据和安全信息，并进行数据格式标准化，为信息共享和数据交换提供数据基础；智能的数据分析，通过信息共享和数据交换对采集的数据进行智能化分析，实现安全信息的集中整理和准确的定损关联，使得技术人员快速的从海量数据中获取有价值的信息；基于专业工作流的安全事件响应机制，为安全事件处理提供合理的流程，并实时监控每个安全事件的发生状态、处理过程和最终结果，是响应安全事件的跟踪器；独立的安全知识管理，提供安全信息发布的平台，包括安全技术交流、安全案例库、系统管理知识、安全维护管理知识、安全新闻等相关信息以及系统补丁

库、常用安全维护工具、工具软件等工具下载，以实现安全知识的共享，提供组织的整体安全水平；

- ***人工智能、大数据与安全技术的结合***

采用了人工智能的机器学习/深度学习技术，基于大数据平台，用海量安全数据进行训练，从而具备检测未知威胁的能力，并有效减少安全运维人员的人工识别工作量。

- ***高效的网络异常行为检测技术***

可识别丰富的网络异常行为模式匹配等检测技术快速鉴别出恶意行为、SSH/FTP 暴力破解、SQL 注入、异常连接、漏洞扫描和漏洞攻击等网络恶意行为。

- ***独特的基因图谱检测技术***

通过结合机器学习、深度学习、图像分析技术，将恶意代码映射为灰度图像，建立卷积神经元网络 CNN 深度学习模型，利用恶意代码家族灰度图像集合训练卷积神经元网络，并建立检测模型，利用检测模型对恶意代码及其变种进行家族检测。基于灰度图像映射的方法可以有效的避免反追踪、反逆向逻辑以及其他常用的代码混淆策略。并且该方法能够有效地检测使用特定封装工具打包（加壳）的恶意代码。

- ***全面的已知、未知威胁检测技术***

通过内置防病毒引擎和威胁情报检测技术对已知威胁进行静态检测；通过基因检测技术对恶意代码的变种进行检测，通过对主机行为和网络行为进行深入分析，对未知威胁进行检测。

全面识别：支持基于指纹和机器学习的资产探测识别，具备网络空间拓扑测绘能力，网络结构分析能力，网络空间资产归属关联能力等。

精准定位：支持挖掘技术和城市级定位能力，全面描述和展示区域内的网络空间信息，为各类应用提供数据和地图可视化的业务支持。

深度挖掘：横向关联漏洞信息、威胁情报信息、图标信息、资产归属信息、自治域信息、DNS 信息等，并对主要协议进行字段深度识别和 AI 数据挖掘。具备对重点威胁组织的持续性主动情报挖掘能力。

持续运营：提供持续的网络空间探测和数据分析能力及工作台；实现漏洞数据、威胁情报数据、其它关联数据的更新和预警；具备独立的 APT 组织发现能力和相关高级威胁的持续追踪溯源能力。

简谈 IDCISP 信息安全管理系统升级改造

陈 敏
（福建省邮电规划设计院有限公司，福建 福州 350003）

摘 要：阐述 IDCISP 信息安全管理系统整体架构、功能结构、系统部署，结合网络安全和数据安全新增功能需求，分析对现有系统升级改造的要求：IDC/ISP 双向网络流量采集，信息安全监测和处置等基础功能，以及网络安全和数据安全的监测和处置等扩展功能。

关键词：SMMS 安全监管系统；ISMS 信息安全管理系统（简称信安系统）；网络安全；数据安全；CU（控制单元）；EU（执行单元）

0 引言

IDC 信息安全管理系统（Information Security Management System，简称 ISMS）是 lDC 经营者建设的具有基础数据管理、访问日志管理、信息安全管理等功能的信息安全管理系统，用于满足电信管理部门和 IDC 经营者信息安全的管理需求。为了有效解决行业网络与数据安全技术监管需求，需要对现有系统做何改造升级呢？本文从 IDCISP 信安系统的架构及功能开始，探讨在网络安全和数据安全新增功能需求下，整个信安系统要做哪些工作来重新构建一个数据安全、网络安全、深度合成信息检测处置等技术能力的安全管理系统。

1 IDCISP 信安系统总体架构

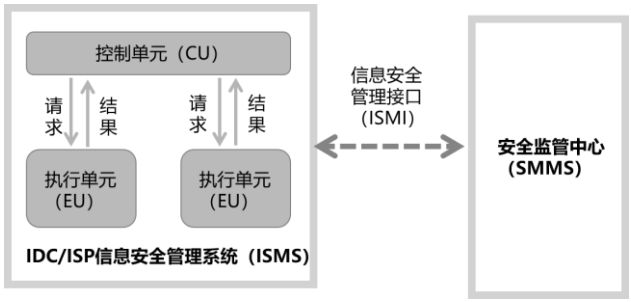


图 1 IDCISP 信息安全管理系统总体架构图

IDCISP 信息安全管理系统总体架构如图 1 所示，包括：控制单元（CU）和执行单元（EU）。

控制单元（CU）：CU 核心控制单元和存储单元集中部署，CU 的核心控制单元负责与安全监管系统（SMMS）进行通信，接收来自 SMMS 的管理指令，并根据要求向 SMMS 上报数据，同时还要实现对各执行单元进行集中管理，完成管理指令的调度、转发和执行及数据的汇总、分析和预警。

执行单元（EU）：EU 部署在各 IDC 机房中，在 IDC 机房出口路由器设备的出口链路上加分光器，经分光器复制之后的一条链路仍连接原有上联的网络设备，另外一条链路接入执行单元的分流设备。分流设备具有一定的过滤功能，将过滤之后的数据传给分析监控服务器，另外也通过一条链路与 IDC 核心网络设备连接，通过发送数据包中断或干扰用户正常业务连接达到对用户流量的控制。

2 需求分析

2.1 必要性

（1）为贯彻落实《中华人民共和国数据安全法》《中华人民共和国网络安全法》，进一步加强数据安全和网络安全技术监管能力建设，统筹构建数据安全监管平台，一体化推进网络安全技术建设。

（2）为满足 IDC 及互联网专线业务的发展需求，满足用户访问 IDC 业务行为分析需求、满足重点 ICP 流量流向分析需求，满足 IDC 业务和流量精细化控制需求，实现对 IDC 业务流量进行安全管控和分析。

（3）积极推进云网融合一体化的建设需求，

实现设备上云。

2.2 功能性

（1）优化信安系统数据采集和处置能力，具备在IDC机房出入口按需采集双向网络流量能力，进一步增强精细化处置能力。

（2）二是扩大系统覆盖范围，满足对专线加密传输流量的数据传输日志分析以及非加密传输流量的按需采集与分析需求。

（3）三是扩展系统功能，新增数据安全模块，支持流量数据识别、数据分级分类，可对数据泄露、跨境流动等行为开展监测溯源和处置;新增网络安全模块，可实现网络攻击、恶意程序、网络异常行为等监测溯源和处置。

3 信安系统改造方案

3.1 总体架构

根据改造功能要求，原有采集执行层EU实现全量流量、特定流量的信安、网安、数安识别、监测、风险发现、处置等功能。而控制存储层CU实现结果日志存储、分析、上报，规则指令接受、转发，采集执行层能力调用。系统升级改造后的总体架构如图2所示：

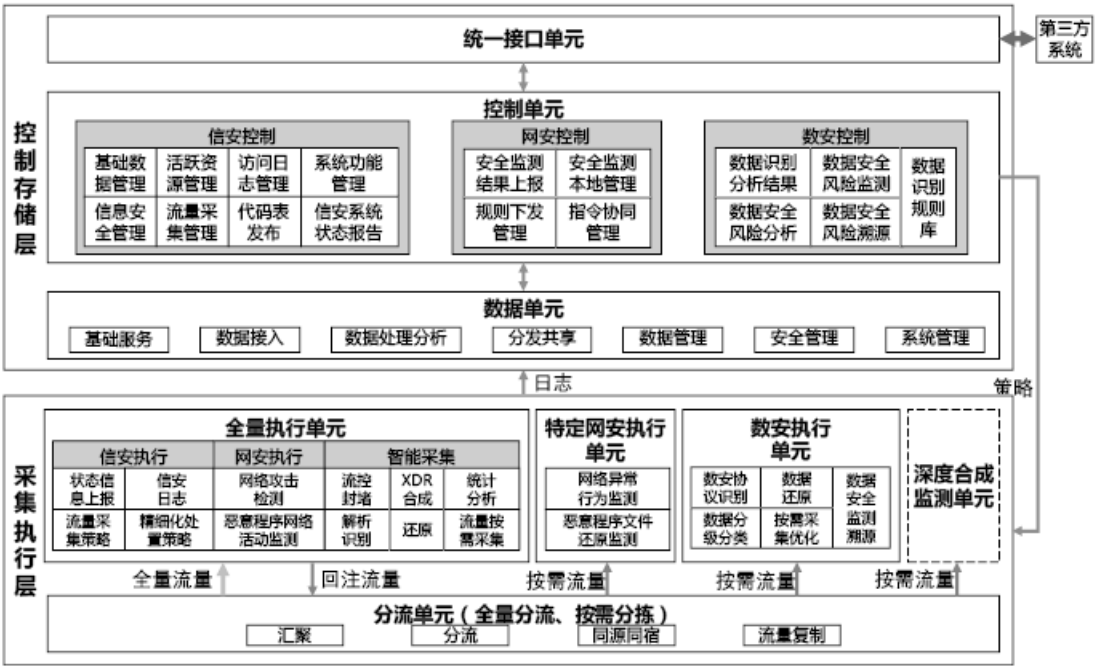


图2 升级改造后的IDCISP总体架构图

3.2 控制单元CU

CU系统优化升级为数据单元、控制单元、统一接口单元几大部分，各部分功能如下：

（1）数据单元：主要包括基础服务、数据接入、数据分析、分发共享、数据管理、安全管理和系统管理等功能，预留数据共享接口。

（2）控制单元：主要包括信安控制、网安控制、数安控制等功能，其中数安和网安为新增功能。

（3）统一接口单元：负责与省管局侧/部侧信安系统以及各控制单元进行统一对接，实现接口统一、规则转化与管理等功能，并向其它第三方系统提供接口。

3.3 采集单元EU

EU系统优化升级为分流单元、全量执行单元、数安执行单元、特定网安执行单元几大部分，各部分功能及实现方式如下：

（1）分流单元：通过部署分流器实现链路汇聚、同源同宿、负载均担等原始流量转发能力。

（2）全量执行单元：通过部署通用基础能力服务器实现统一DPI服务器（EU服务器）功能，具备网络攻击、恶意程序网络活动监测等网安功能，访问日志精细化处置能力，实现按需流量采集等信

安功能。

（3）数安执行单元：通过部署数安扩展能力服务器实现数据监测识别、还原、分级、分类、安全风险监测等数安功能。

（4）特定网安执行单元：通过部署网安扩展能力服务器实现恶意文件还原、网络异常行为监测等网安功能。

4 系统部署

系统流量采集覆盖范围为 IDC 出口双向流量、互联网专线出口链路双向流量。

IDCISP 信安系统部署如图 3 所示，建议如下：

1）控制单元（含统一接口单元、数据单元）统一部署于省中心；有条件的可以采用存算分离的方式进行部署，即接口单元和控制单元统一部署，数据存储单元分散部署；

2）分流单元、全量执行单元、特定网安执行单元、数安执行单元主要部署于各地市 IDC 机房或汇聚机房；

3）全量执行单元接收和处理全量流量，特定网安执行单元、数安执行单元接收和处理按需流量。

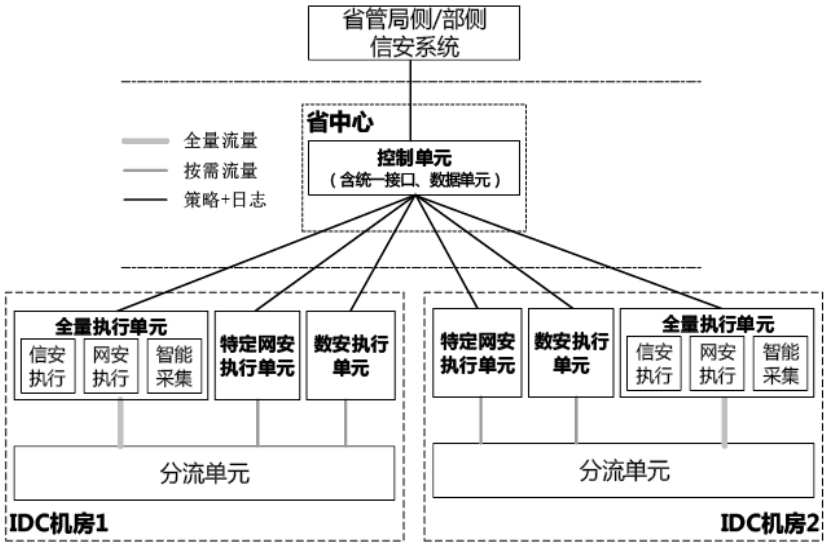


图 3 IDCISP 信安系统部署示意图

5 系统云化

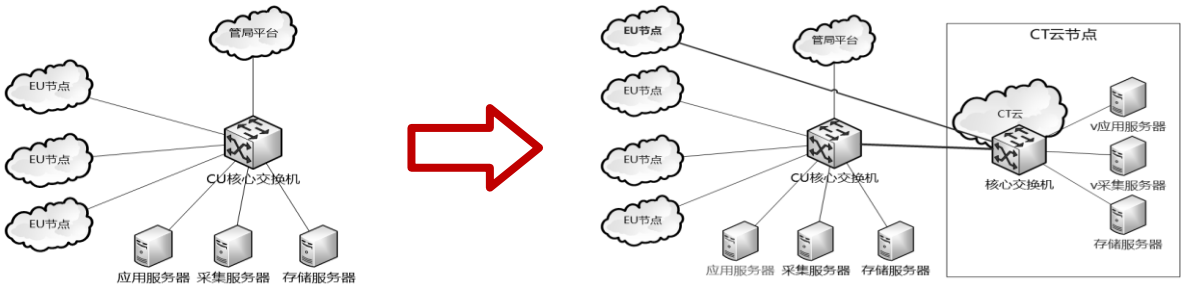


图 4 IDCISP 信息安全管理系统云化示意图

IDCISP 信息安全管理系统主要采用的是 X86 架构服务器，且各网元接口均已实现 IP 化，具备虚拟条件。其中 CU 控制单元根据其网络架构特征及所用设备的特性，可以采用虚拟化技术实现云化部署。具体云化方案如图 4 所示，云化部署建议：

（1） 在云节点新建 CU 节点；

（2） 应用服务器由云资源池提供虚拟资源，便于部署新 CU；

（3） 采集服务器、存储服务器结合现有资源，按需扩容，由云资源池提供虚拟资源或物理设备；

（4） 云资源池节点与原CU节点通过传输专线组网，采集服务器和存储服务器不同节点部署时，尽量匹配资源，减少资源消耗；

（5） 新建EU节点尽量回传新CU所在云节点。

6 结束语

本文介绍IDCISP信息安全管理系统并对系统的升级改造进 行简单描述，其中的信息安全、网络安全和数据安全等与我们的生活息息相关。IDCISP信安系统针对IDC及互联网专线，通过各种技术手段，为依法加强互联网管理，保障信息、数据及网络安全，营造绿色、健康、有序的互联网环境，净化网络不良内容，提升网络服务品质，促进互联网文化的繁荣发展和维护社会稳定提供有效的技术手段和管理支撑。

参考文献:

[1]《如何构建 IDC/ISP 信息安全技术管理系统》，作者：马昌军；张玲，通信世界

[2]《关键信息基础设施安全保护条例》《数据安全法》和网络安全等级保护制度解读和实施，作者：郭启全等，电子工业出版社

陈敏（1977—），男，福建福州，本科毕业于北京邮电大学信息管理与信息系统专业，通信工程师，工作单位为福建省邮电规划设计院有限公司，主要从事城域网、IDC网络、网络和信息安全等专业的研究工作。

TextRCNN 模型结合联邦学习的非结构化数据分类分级研究

郑 炎
（中电福富信息科技有限公司，福建 福州 350001）

摘 要：本文给出了一种使用深度学习技术进行非结构化数据的分类分级识别方法，用于降低传统方案中对脚本规则和正则的依赖。从而大幅提高模型的准确率以及在面对未知特征数据时的鲁棒性。并结合当下数据分类分级在实际生产环境中遇到的信息孤岛以及用户数据敏感性问题，提出了一套基于联邦学习的训练方案。

关键词：数据分类分级；联邦学习；深度学习；TextRCNN；机器学习

0 背景

随着国家在数字化转型的大力推进，政府、运营商以及企业的数据量增大,数据流向越来越复杂，安全风险加大。为了保障数据安全，就是要使数据持续处于有效保护和合法利用状态。

根据 Risk Based Security 的最新报告，2021 年第一季度数据泄露的数据量猛增至 84 亿，与 2020 年第一季度相比增长了 273%，创下至少自 2005 年详细报告开始以来的同期记录。

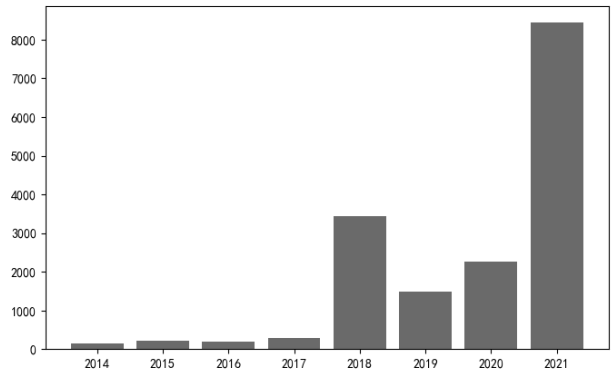


图 1 每年第一季度报告的数据泄露数量(百万)

2021 年 6 月 10 日,《中华人民共和国数据安全法》已由中华人民共和国第十三届全国人民代表大会常务委员会第二十九次会议正式通过，并于 2021 年 9 月 1 日正式施行。根据《数据安全法》第二十一条，国家建立数据分类分级保护制度，根据数据在经济社会发展中的重要程度,以及一旦遭到篡改、破坏或者非法获取、非法利用，对国家安全、公共利益或者公民、组织合法权益造成的危害程度，对数据实行分类分级保护。数据分类分级作为数据安全的第一步，在数据安全治理过程中至关重要，一般企业的数据可以分为公开数据、非公开数据（敏感数据）。为此需要把主要精力放在敏感数据的管控上，制定精细化的管控原则。根据不同数据级别，实现不同的安全防护，避免敏感数据泄露给企业造成重大损失。各行各业的分类分级标准虽然如雨后春笋般大量颁布，由于缺少辅助工具和持续优化的运营经验，面对规范要求中繁杂的数据类型往往无从下手。再有规范中一般分级规则是各监管单位、核心企业根据业务模型讨论出来的理想情况。在实际落地过程中发现，数据的分类分级既需要熟知行业业务的业务型人才，也需要具备能够从数据内容形态上抽象出分类规则的专业人员。正是由于这些原因，业界急需要一套能够实现自动化并能够贴合行业的数据分级分类的解决方案。

1 挑战

在自动化数据分类分级上的问题上，现在业界

常用的解决方案是采用人工编写规则加机器学习辅助的方式。具体步骤如下：

• 先收集获取目标行业的各种类型的数据样本，样本要尽可能覆盖该行业大部分的业务情况。

• 人工对所有的数据进行清洗和筛选，过滤掉无用的数据样本。

• 专业人士依据行业制定的分类分级规范，结合数据样本，编写合适的规则（规则包括了正则表达式、规则脚本等）

• 对于一些特征形式比较难确定，很难用编写规则来进行匹配的数据样本，则通过训练机器学习模型来进行分类。

传统的这种方案在数据分类分级这个问题上存在着许多问题：

1. 规则的编写需要有很强的专业能力，这里面包括对规则的编写能力以及发现数据特征的敏锐直觉，而这种能力往往需要多年的经验积累，拥有较高的专业门槛。

2. 由于样本数据需要尽可能覆盖行业大部分的业务，所以最好的贴合业务的样本数据就是行业生产环境的数据。而这些数据中往往带有许多行业内的敏感信息，即便数据在交付给专业规则编写人员前，一般都会进行数据脱敏处理。但是数据脱敏又会带来额外的问题，一是脱敏的时候可能会有遗漏，导致敏感信息还是泄露出去，二是脱敏后的数据往往是一些假数据，这些假数据往往自相矛盾，导致数据与数据之间的关联特征丢失。

3. 遍历规则匹配性能低下，规则匹配的方式需要将输入数据与编写的规则进行逐一匹配，规则数量不大的时候，对性能影响不明显，但是往往实际生产环境的分类规则都十分庞大，这就导致了分类性能往往达不到用户的要求。

4. 机器学习作为分类分级的最后辅助手段，往往给到的数据样本不够充分，模型的拟合效果不理想，导致训练出的模型准确率不高。

也正是由于上面种种原因导致数据分类分级在实际落地过程中往往无法达到预期，这也正是本文给出的方案所需要解决的问题。

2 方案

传统解决方案中，大部分的工作量集中在规则的编写过程。其实人工的规则编写到结果输出从本质上的理解可以分为以下几个步骤：

1. 样本数据的整理。
2. 数据特征的提取。
3. 分类规则的编写。
4. 输入实际数据匹配规则。
5. 输出结果。

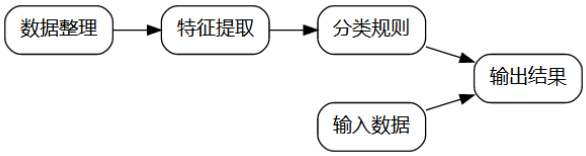


图 2 人工编写规则的具体流程

在这过程中，从数据整理、特征提取再到分类规则都是由有经验的专业人员完成。按照目前行业平均三级分类类别数量在 1000 以上，每种类别下需要应对的数据样本模板一般在 10 种以上可能性，光是一个行业就需要编写上万条规则。而且每条规则需要进行大量的测试验证，考虑各种边界情况才能投入使用。这样一来，一个真正能够投入使用的数据分类分级产品，需要的人工成本是巨大的。同时一条数据需要匹配上万条规则，对产品的性能影响也是巨大的。

另一方面，对于人工无法编写规则来进行匹配的数据，一些数据分类分级产品采用了机器学习模型辅助的方案。具体做法是人工先将数据分类好后提交给平台，由平台提供机器学习算法如朴素贝叶斯、决策树、条件随机场、隐马尔可夫等模型进行训练学习。而这个过程中，一方面需要将样本数据上传汇总到平台，这个过程中就很有可能有数据泄露的风险。另一方面，由于集中式学习对算力的要求和数据由于规则筛选后的数量大幅减少，往往只能使用一些简单的模型进行分类，从而导致模型分类的效果很难达到预期。

为了解决传统方案的人工成本问题，本文的方案以深度学习作为问题的解决切入点。

引入深度学习的第一大优势是表征学习，通过多层神经网络来对样本数据进行拟合，由神经网络

的监督式学习方式来自我学习得到特征提取的能力。从而实现端到端的训练过程，去掉了人工在特征提取上的工作。第二大优势是分类能力，相较于人工编写规则的逐条匹配，神经网络本身的全连接层的分类是一次计算后直接得到各个分类的概率向量，在性能上明显强于规则匹配。另一方面，敏感数据出于反泄露的安全要求，必然会形成数据孤岛。故而，传统的各种方案必然要面临样本数据匮乏的瓶颈。为了解决数据孤岛问题，本文引入了联邦学习的方式，将联邦学习和深度学习技术进行融合。

采用深度学习结合联邦学习的方案主要的步骤如下：

1. 数据样本分类。
2. 下载模型结构以及初始权重
3. 训练模型。
4. 上传权重梯度。
5. 汇总更新模型权重。
6. 输入生产数据。
7. 输出结果。

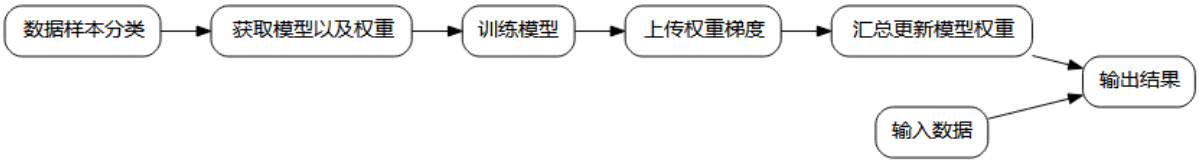


图 3 深度学习结合联邦学习的工作流程

- **样本数据的分类分级**

前期这块工作内容可以由业务系统使用人按照目标行业的分类分级标准对实际生产环境的数据进行分类分级作为样本数据。

- **获取模型以及权重**

业务使用者可以从联邦学习的统一平台下载初始模型以及必要的环境和训练脚本，然后将分类分级后的样本数据输入模型进行训练，再配合神经网络的反向传播来进行权值更新，从而使模型具备了自我学习的能力。模型采用 TextRCNN 结构，整个模型的模型结构如下图 4 所示：

- **Word Embedding 层**

Word Embedding 层的主要作用是对文本信息中的词进行空间映射，使得词具有空间语义。也就是将原来表示词的 One-Hot 向量（一种词与词之间完全正交的向量表示方式）通过 Word Embedding 后，向量的维度会得到压缩，同时语义相近或者分类相近的词，在空间距离上也会相近。Word Embedding 层的数学运算方式：

$$C = W * X (X \in R^n, C \in R^m, W \in R^{m \times n}, m \ll n)$$

其中X为输入的 One-Hot 编码向量，W为 Word Embedding 矩阵，C为 Embedding 后输出的词向量。

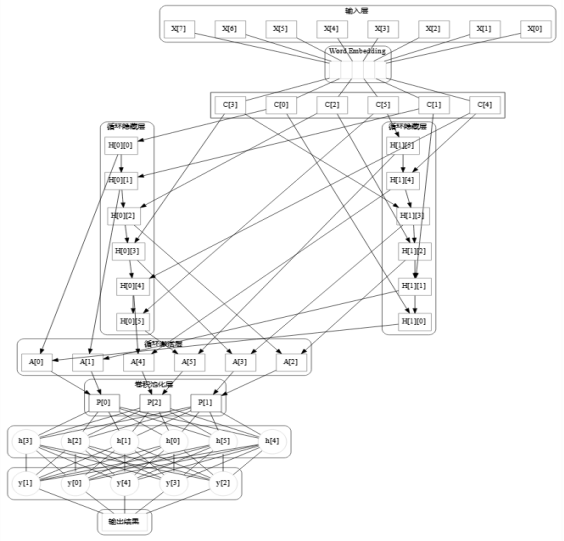


图 4 模型结构（该图只是用来理解模型的结构，不代表真实的维度信息，具体的维度信息需要根据业务场景进行设计）

- **双向循环神经网络层**

经过 Word Embedding 层后的词已经具有词相关的语义信息，但是为了提取上下文关联关系的特征信息，本方案采用了循环神经网络层对上下文关联信息进行特征提取，循环神经网络层采用的是双向 GRU 实现，其数学表示：

$$a_t = \sigma(W_a * [\ ''_{t-1}, x_t])$$
$$r_t = \sigma(W_r * [\ ''_{t-1}, x_t])$$
$$\tilde{''}_t = tan\ ''(W * [r_t *\ ''_{t-1}, x_t])$$
$$''_t = (1 - a_t) *\ ''_{t-1} + a_t * \tilde{''}_t$$

其中t代表t时刻的状态，W代表 GRU 的隐藏层的神经元，a为更新门控，r为重置门控，h用于记录该时刻的特征信息。

σ是 Sigmoid 函数，函数的表达式如下：

$$\sigma(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

tanh 函数的表达式如下：

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

- **卷积池化层**

卷积池化层的作用是用于最后的特征提取以及特征增强。其原理是通过三个通道数为 1 的一维卷积核对前面的循环神经网络层的输出进行卷积操作，最后得到三个通道的输出。使用卷积操作的是因为卷积比起多层感知机来说可以使用更少的权重参数，同时又兼顾有很好的特征提取能力。卷积后的特征先分别经过 Leaky ReLU 激活函数后再进行最大池化操作来从而对数据维度进行压缩并增强其中的特征。

这部分用数学公式表示：

$$Z^{[i]} = (A * W)(j,k) = \sum_m \sum_n (A(m,n)W(j-m, k-n))$$
$$A^{[i]} = LeakyRelu_\lambda(Z^{[i]})$$
$$P^{[i]} = MaxPool(Z^{[i]})$$

其中W为卷积核，A为前面的循环神经网络层的输出。

采用 LeaKy Relu 函数作为激活函数的原因是考虑到 Relu 会降低梯度消失的可能性：

$$LeakyRelu_\lambda(z) = max(\lambda z, z) = \begin{cases} z, & if\ z \geq 0 \\ \lambda z, & if\ z < 0 \end{cases}$$

其中λ为模型的超参数。

- **全连接分类层**

接下去再将分别对卷积后三个通道的输出在第一维度上进行堆叠，这样可以将：

$$M = P^{[1]} \oplus P^{[2]} \oplus P^{[3]}$$

由于后面要进行全连接，所以要将拼接后的结果进行展平成向量：

$$F = Flatten(M)$$

展平后接下来就是常见的两层全连接层，第一层还是采用 Leaky Relu 激活函数，最后的输出层采用 Softmax 进行分类输出：

$$Z^{[4]} = (W^{[4]})^T * F + b^{[4]}$$
$$A^{[4]} = Relu(Z^{[4]})$$
$$Z^{[5]} = (W^{[5]})^T * A^{[4]} + b^{[5]}$$
$$\hat{Y} = Softmax(Z^{[5]})$$

Softmax 函数是用来对传入的向量参数的每一维的值都分别求做 e 的指数，然后再分别除以所有值 e 的指数的和：

$$Softmax(V) = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

- **联邦学习训练过程**

前向传播后，通过反向传播然后用梯度下降法来训练模型的权重。首先可以拿预测值与真实值去做比较，采用交叉熵作为损失函数：

$$\xi(\hat{Y}, Y) = -\sum_{j=1}^{n} Y_j\ log\widehat{Y_j}$$

n的值等于输出分类的个数，$\hat{Y}$是预测值Y是真实值。

如果是批量训练数据的话还需要对批量数据集计算平均损失函数，那么损失函数的表示形式就变成下面这样：

$$J(\hat{Y}, Y) = \frac{1}{m} * \sum_{i=1}^{m} \xi(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{m} * \sum_{i=1}^{m} (-\sum_{j=1}^{n} y_j^{(i)}\ log\hat{y}_j^{(i)})$$

上面的Y代表这一批的真实值，$\hat{Y}$代表这一批的预测值，y和$\hat{y}$分别代表这一批里的单条数据。

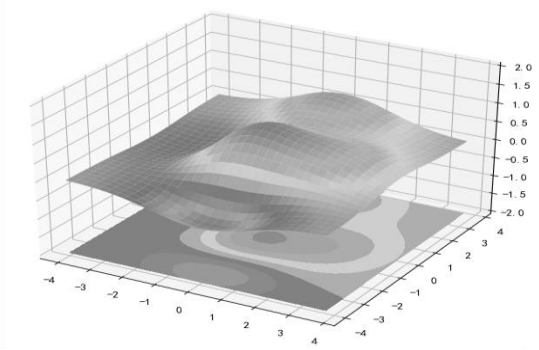


图 5 损失函数 3 维可视化示意图

接下去采用链式求导法则对所有待训练的模型参数进行梯度计算。如图 5 所示的损失函数的 3 维示意图，采用梯度下降法是通过计算当前点在损失函数的梯度，然后将权值参数按照梯度方向的反方向进行更新。而不同于传统的模型训练，联邦学习需要将这些权值参数梯度信息进行记录，然后最终汇总更新模型参数的时候将所有孤岛上计算出的梯度数据用于更新最终的模型。

权值偏导的计算方式：

$$J(\hat{Y},Y)=J(A^{[k]},Y)=\frac{1}{m}*\sum_{i=1}^{m}\xi(\hat{y}^{(i)},y^{(i)})$$

$$\nabla A^{[k]}=\frac{\partial J(A^{[k]},Y)}{\partial A^{[k]}}=\frac{\partial J}{\partial A^{[k]}}=J'\quad(A^{[k]},Y)$$

$$\nabla Z^{[k]}=\frac{\partial J}{\partial Z^{[k]}}=\nabla A^{[k]}*g^{[k]\prime}\quad(Z^{[k]})=\frac{\partial J(A^{[k]},Y)}{\partial A^{[k]}}*g^{[k]\prime}\quad(Z^{[k]})$$

$$\nabla W^{[k]}=\frac{\partial J}{\partial W^{[k]}}=\nabla Z^{[k]}*A^{[k-1]}$$

$$\nabla b^{[k]}=\frac{\partial J}{\partial b^{[k]}}=\nabla Z^{[k]}$$

$$\nabla A^{[k-1]}=\frac{\partial J}{\partial A^{[k-1]}}=\nabla W^{[k]}*\nabla Z^{[k]}$$

$$\nabla Z^{[k-1]}=\frac{\partial J}{\partial Z^{[k-1]}}=\nabla A^{[k-1]}*g^{[k-1]\prime}\quad(Z^{[k-1]})$$

$$\nabla W^{[k-1]}=\frac{\partial J}{\partial W^{[k-1]}}=\nabla Z^{[k-1]}*A^{[k-2]}$$

$$\nabla b^{[l-1]}=\frac{\partial J}{\partial b^{[k-1]}}=\nabla Z^{[k-1]}$$

...

$$\nabla A^{[1]}=\frac{\partial J}{\partial A^{[1]}}=\nabla W^{[2]}*\nabla Z^{[2]}$$

$$\nabla Z^{[1]}=\frac{\partial J}{\partial Z^{[1]}}=\nabla A^{[1]}*g^{[1]\prime}\quad(Z^{[1]})$$

$$\nabla W^{[1]}=\frac{\partial J}{\partial W^{[1]}}=\nabla Z^{[1]}*X$$

$$\nabla b^{[1]}=\frac{\partial J}{\partial b^{[1]}}=\nabla Z^{[1]}$$

其中J代表当前批次的平均损失，k代表神经网络的层数，m代表当前批次的样本数量，A代表前向传播各层经过激活函数后的输出值，g代表各层的激活函数，Z代表各层经过激活函数前的输出，W代表各层的权值，b代表各层的偏置值，梯度前面用∇表示，例如W权重的梯度用∇W表示。

将各层的梯度信息进行保存到特定文件中，保存的信息包含了：($\nabla W^{[1]},\nabla b^{[1]},\nabla W^{[2]},\nabla b^{[2]},\ldots,\nabla W^{[k]},\nabla b^{[k]}$)。

- **汇总更新模型权重**

联邦学习的最大优势就是不需要将样本数据上传到总模型训练平台，可以由各自孤岛进行训练后把梯度信息上报汇总即可。这样的好处是很好的规避了敏感数据的泄露以及实现了分布式训练的效果，极大的加快了训练的速度。

最终模型的权重更新只需要用每个孤岛上报上来的梯度信息来对各层的权值更新：

$$\boldsymbol{for}\quad(\boldsymbol{i=l\ to\ k}):$$
$$\boldsymbol{W^{[i]}:=W^{[i]}-\mu*\nabla W^{[i]}}$$
$$\boldsymbol{b^{[i]}:=b^{[i]}-\mu*\nabla b^{[i]}}$$

μ代表学习率。

- **结果输出**

通过 Softmax 输出的结果为结果的概率分布，即如果定义的分类类型为 5 类，那么输出的结果就为一个 5 维的向量，向量的五个维度的值的和为 1。值最大的那个维度对应的分类就是模型预测出的分类结果。

3 总结

我们团队测试了 7 组不同样本数量采取 TextRCNN 模型与人工规则匹配法进行对照实验，结果如图 6 所示。可以看到在样本数量只有 5000 以下的时候，规则匹配的效果比较好，但是随着样本数量的增加，规则匹配由于受限于人工自身的上限以及人工的时间成本，所以很难有更大的提升空间。当样本数量超过 10000 的时候，TextRCNN 的准确率已经远远超过了人工规则匹配的方式。由此可见，深度学习在分类分级任务上有着人工所不能比拟的优势。

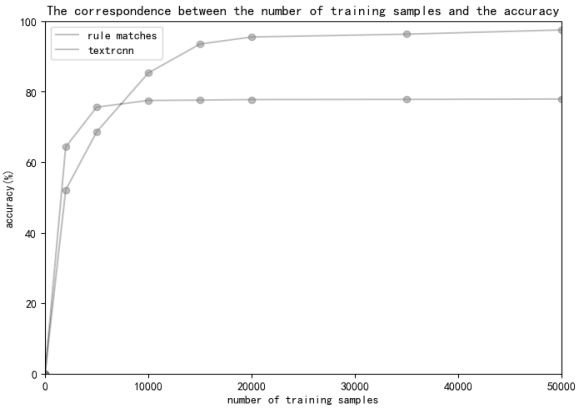


图 6 对比人工规则匹配与 TextRCNN 的样本数量与预测准确率的对应关系

通过采用 TextRCNN 模型的方式，大大降低了人工编写规则的人工成本。同时极大程度的降低了分类分级工作的门槛，使得了解行业业务和规范

的非专家人士也可以很好的参与分类分级的实际工作中，大大提升分类分级的工作效率。另一方面，联邦学习的方式，使得各个业务系统不需要将数据泄漏给算法平台进行训练，可以各自在各自的业务平台上将模型进行预训练后，将权重数据上传给算法平台进行模型的汇总更新。这样可以有效的保证业务系统的数据安全和训练效率。

参考文献:

[1] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.

[2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors[J]. Nature, vol. 323, no. 6088, pp. 533 - 536, 1986.

[3] A Communication-Efficient Federated Learning Method with Periodic Averaging and Quantization, Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, 2020

[4] Recurrent Convolutional Neural Networks for Text Classification, 2015

作者简介

郑炎，男，现就职于中电福富信息科技有限公司，福富大学星级讲师，算法工程师，主要研究方向为机器学习、深度学习、系统架构以及网络和信息安全。联系电话：13405999830；邮箱：zhengyan2@ffcs.cn