网络空间信息系统模型与应用

王继龙[1,2,3]，庄姝颖[1,3]，缪葱葱[2,3]，安常青[1,3]
（1. 清华大学网络科学与网络空间研究院，北京 100084；2. 清华大学计算机科学与技术系，北京 100084；
3. 清华大学北京信息科学与技术国家研究中心，北京 100084）

摘 要：针对网络空间可视化往往还在使用传统的地理信息系统，难以真正展现网络空间特点和规律的问题，提出了多尺度、多维度、多视图的网络空间信息系统模型，将网络空间信息系统定位为与地理信息系统相平行的概念，即地理信息系统支持以地理坐标系为基础的可视化表达，网络空间信息系统支持基于网络空间自身坐标系的可视化表达。同时定义了网络空间信息系统的基本概念模型，为之后的研究奠定理论基础，并阐述了网络空间信息系统关键技术，以及部分关于坐标系、比例尺、多维度表达模型等探索性研究工作和应用案例。研发的原型效果表明，网络空间信息系统更便于网络空间事务的展现和处理。

关键词：网络空间；网络空间坐标系；网络空间地图；网络空间信息系统
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2020018

Model and application of cyberspace information system

WANG Jilong[1,2,3], ZHUANG Shuying[1,3], MIAO Congcong[2,3], AN Changqing[1,3]
1. Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100084, China
2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
3. Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

Abstract: For the visualization of cyberspace is still using the traditional geographic information system, which makes it difficult to truly express the characteristics and laws of cyberspace, a multi-scale, multi-dimensional and multi-view cyberspace information system model was proposed to take the cyberspace information system as a concept parallel to the geographic information system, that was, the geographic information system supported visualization based on the geographic coordinate system, and the cyberspace information system supported visualization based on the coordinate system of cyberspace itself. Then, the basic concepts of cyberspace information system were defined to provide a theoretical foundation for later research, and the key technologies, part of exploratory research work including the coordinate system, scale, multi-dimensional expression model and application cases were expounded. According to the prototype developed, the cyberspace information system is more convenient for the display and processing of cyberspace affairs.

Key words: cyberspace, cyberspace coordinate system, cyberspace map, cyberspace information system

1 引言

随着互联网的普及以及信息技术的发展，网络逐渐空间化，网络空间作为人类新的疆域已取得广泛共识，如何更好地探索和表达网络空间受到各国的广泛关注。2012 年起，美国相继启动“藏宝图计划”[1]和“X 计划”[2]，用于研究实时、交互的全球互联网地图，为网络空间信息的研究提供了基础支撑。2014 年，俄罗斯卡巴斯基实验室[3]发布网络威胁地图，致力于网络活动情况的实时表达。2018 年，中国互联网安全大会（CISC, China Internet Security Conference）中展出了智慧城市系统，对网

收稿日期：2019–09–17；修回日期：2019–12–10
基金项目：国家重点研发计划基金资助项目（No.2016YFB0801301, No.2016QY12Z2103）
Foundation Item: The National Key Research and Development Program of China (No.2016YFB0801301, No.2016QY12Z2103)

络基础信息进行整合后集中显示，实时监测并及时响应网络安全攻击。

然而，目前网络空间测量和态势感知工作所面临的一个共性问题是缺乏网络空间自身的坐标系和地图支持，往往还在使用传统的地理信息系统，难以真正展现网络空间的特点和规律。

网络空间作为平行于物理空间的新世界，至今尚未建立真正意义上的空间模型。物理空间中有相应的空间坐标系、地图、地理信息系统作为信息表达基础；网络空间突破了传统物理空间的时空限制，使传统物理空间特征在网络空间的重要性降低，因此急需建立网络空间自身的空间模型和空间信息系统。

基于之前关于网络空间坐标系和网络空间地图的研究工作[4]，本文提出了多尺度、多维度、多视图的网络空间信息系统模型。具体贡献如下。

1) 应用地理信息系统的方法论进行交叉学科研究，初步对网络空间的空间特性和网络空间信息系统进行系统化的定义和描述，分析网络空间不同于地理空间的特性和规律，建立网络空间自身的空间模型和空间信息系统的难点和挑战，为网络空间信息系统的研究奠定理论基础。

2) 针对传统网络空间表达模型的缺陷，本文在恒定网络空间坐标系设计的基础上，定量分析了 Hilbert 曲线映射算法在网络空间的层次性、区域性表达，以满足用户对不同观察粒度下网络空间要素的可视化需求。

3) 建立网络空间信息系统模型。本文围绕着如何从多尺度、多维度、多视图的角度表达虚拟网络空间这一关键问题，设计网络空间比例尺，用于多尺度表达复杂多样的空间信息；构建 2 种多维度表达模型，基于网络空间信息系统可实现坐标维度扩展的表达方法这一重要特征，支持不同场景下网络空间的多维度信息表达；结合网络空间地图和物理空间地图，提供多视图全面观察网络空间的方法。

4) 研发网络空间信息系统原型，将其应用于网络空间资源要素的定位和表达、流量监控与故障管理以及网络安全攻击可视化等场景，取得良好可视化效果。

2　网络空间的空间特性

网络空间现已成为继陆、海、空、天的第五大疆域，目前尚未形成统一的基本空间概念模型，导致网络空间的表达研究仍缺乏基础共识。本节借鉴地理信息系统的探索思路，对网络空间的空间特性和网络空间信息系统进行初步的系统化描述和定义，并分析后续研究面临的关键技术难点和挑战，为网络空间信息系统的研究奠定理论基础。

1) 网络空间特性

网络空间作为独立存在的多维度虚拟空间[5]，具有以下特性。

① 距离无关性。不同于地理空间中距离的概念，网络空间的信息传输本质以及通信设备导致的时空压缩，使其任意节点间的距离体现在瞬时的网络时延中，与物理距离无关，突破了传统时空限制，构成一个无传统距离概念的空间。

② 复杂性。网络空间由复杂多样的异质网元构成，既包括实体资源如服务器、交换机、链路、终端节点等基础设施，也包含应用服务、账号、直接或间接连接互联网的系统、网络数据等虚拟资源。在不同网络层次、观察粒度下，表现出不同的空间资源要素。

③ 变化性。网络空间中海量信息自由流动，导致网络空间瞬时多样。同时，网络用户既可以成为信息发布者，也可以成为信息接收者，其行为的多样性直接导致沟通对象动态变化。因此，兼具稳定性和实时性将成为网络空间的表达基础。

④ 区域性。网络空间的网络特性，使其超越了传统意义上的地理概念[6]，形成了新的空间区域形式，包括局域网、部分连续的 IP（Internet protocol）地址段等，对外表现出明显的网络特征（路由、流量、用户行为等）相似性。

正因为上述网络空间的空间特性，基于地理信息系统表达网络空间难以体现其独有的特点和规律。地理空间中，表达模型通过投影映射将三维球面转换成二维平面，以经度和纬度作为基础向量构建二维地理坐标系，同时提供了比例尺等概念用于明确地理空间尺度[7]，促进了地理空间统一绘制背版的形成。网络空间弱化了距离等概念，类似的简单直观模型并不存在。如何借鉴地理空间学的相关理论，建立网络空间自身的空间模型和空间信息系统将成为研究的热点和难点。

2) 网络空间信息系统

地理信息系统依赖于计算机的存储功能。将实际生活中的物理信息存储在地理信息系统数据库

中，使人们能将实际生活中的空间信息模拟化[8]。类似地，网络空间信息系统的构建则主要考虑如何基于网络测量数据实现虚拟网络空间信息的可视化表达。具体研究主要包含以下 3 个方面。

① 网络空间坐标系。网络空间坐标系作为网络空间信息表达的重要参照基础，至今尚未形成统一的坐标向量、空间维度定义。考虑到网络空间复杂多变的空间特性，本文试图回答如何选取类似经纬度的恒定坐标向量维度；如何设计坐标系映射方法，满足能够控制坐标系统参数实现网络信息层次化表达的需求；如何反映网络空间的区域特性等基础性问题，完成网络空间自身坐标系的构建。

② 网络空间信息系统模型。网络空间数据量大且维度多样，如何从多尺度、多维度、多视图的角度表达虚拟网络空间，将成为网络空间信息系统模型研究的关键。本文借鉴地理信息系统模型研究的方法论，首先，设计网络空间比例尺，用于伸缩表示复杂网络资源要素的层次结构，符合人类由远及近认知网络空间信息的思维。其次，通过叠加图层或者添加新的空间坐标维度的方式，构建特定主题的网络空间专题地图，进而实现网络空间多维信息表达。最后，研究网络空间与物理空间的映射方法，完成数据的筛选整合以及多视图表达，构建网络空间信息系统。

③ 应用场景可视化。为了满足用户基于网络空间信息系统的多方面需求，体现网络空间信息系统的重要作用，本文以网络空间要素定位和表达、网络空间监控与管理以及网络空间安全等场景为例，实现网络空间深层次剖析和可视化。

3 传统网络空间表达模型

近年来，有关网络空间可视化表达的研究主要集中于 2 种模型。一种是基于地理信息系统的网络空间可视化模型，即基于地理坐标系映射网络空间，侧重于表达网络空间要素的地理属性特征；另一种是基于网络拓扑的网络空间可视化模型，以拓扑学理论为基础，基于单元与连接线等描述形式表达网络空间的拓扑连接关系，并采用空间剖分降维的方式描述子单元拓扑信息。这 2 种模型对于网络空间信息系统的建立均提供了一定的借鉴和指导意义，但均难以从本源角度描述网络空间的空间特性及规律。

3.1 基于地理信息系统的网络空间可视化模型

基于地理信息系统的网络空间可视化模型关注虚拟网络空间与物理空间的融合。文献[9-10]最早开启基于地理信息系统的网络空间可视化模型的研究，从传统地理学领域的研究核心“空间”与“位置”出发，定义网络空间是由信息网络连接而成的计算机空间的集合，网络位置则是网络基础载体与传统地理空间基础设施之间的映射，通过该映射关系实现网络空间与物理空间之间的关联。地理网络空间模型以地理坐标系为基础，其绘制方法可借鉴传统的地理地图学原理，是目前主流的网络空间描述和表达方案，被工业界和学术界广泛应用。

纵观目前的研究，基于地理信息系统的网络空间可视化模型主要从网络基础载体、性能参数以及应用安全等角度描述网络空间。文献[11]研究 DNS（domain name system）泛播服务载体的地理分布特征，通过匹配客户端对应的 DNS 服务载体构建映射关系，精确绘制 DNS 服务载体的真实地理坐标，进而协助研究网络空间 DNS 服务的负载均衡问题。Huffakr 等[12]将网络空间进行空间划分，并将每个空间映射到相应地理区域，在此基础上，每个机构通过管辖该区域的基础载体，即可实现网络空间区域的自治系统管理。此外，CAIDA（center for applied internet data analysis）还绘制了网络空间流量流向在地理空间的分布图[13]，以及 Witty 蠕虫宿主主机的地理空间分布与时空传播模型图[14]，有助于指导地理基础设施部署和地理安全防范。

基于地理信息系统的网络空间可视化模型基于地理坐标系映射网络空间，表达网络空间要素的地理属性特征，协助从地理空间层面完成网络空间的管理。网络空间依托于地理实体存在，因此网络空间的时空关系可以在地理空间中体现。但是网络空间作为独立的虚拟新空间，突破了传统物理空间的时空限制，地理距离的意义极大淡化，且具有独特的网络区域性以及层次结构，使传统地理空间特征在网络空间的重要性降低。因此，基于地理信息系统的网络空间可视化模型无法真正揭示网络空间的本源空间特性。

3.2 基于网络拓扑的网络空间可视化模型

另一部分研究学者认为，对于网络空间而言，其物理距离的概念正逐渐弱化，而拓扑连接关系对于认知网络空间显得尤其重要。因此，他们利用网络单元（V）和链路（E）将网络空间抽象成网络拓

扑关系图 $G(V,E)$，进而构建基于网络拓扑的网络空间可视化模型，将虚拟的网络空间抽象成多层次的拓扑网络结构，完成网络空间与拓扑空间之间的映射。其中网络拓扑的构建与绘制可借鉴传统拓扑学原理。

拓扑网络空间模型的研究主要围绕自治系统（AS, autonomous system）层、路由器层和 IP 层，实现多层次的网络空间描述和表达。AS 层拓扑模型以 AS 为基础单元，将网络空间进行粗粒度的空间划分，以反映整个网络空间要素的连接关系。Mahadevan 等[15]结合边界网管协议（BGP, border gateway protocol）数据和 Traceroute 测量数据等多元数据集，提出基于联合概率分布精确绘制网络空间 AS 拓扑连接关系图的方法。路由器层拓扑模型专注描述网络空间的局部连接特征。Keys 等[16]利用研发的别名解析系统，构建网络空间局部区域拓扑图，实现网络空间子区域管理。IP 层拓扑模型则通过探测链路层的连接情况，支持主机和接口级别的网络空间拓扑信息可视化，文献[17]基于 SNMP（simple network management protocol）构建实体单元的拓扑图，有助于网络空间实体单元和连接链路的管理。

综上所述，拓扑网络空间模型基于拓扑关系能较好地映射网络空间，其中每个网络空间单元可以由多个网络空间子单元构成，进而将网络空间进行剖分降维，实现对网络空间多层次、细粒度的认知。但拓扑网络空间可视化模型存在一定的本质问题，考虑到网络空间瞬时多样的特性，随着网络中新节点的加入与旧节点的离开，可能导致拓扑节点时刻动态变化，该模型无法提供恒定的描述和表达网络空间的方法，更不满足坐标系自身恒定不变的要求。

4　网络空间坐标系

对于网络空间而言，上述基于地理信息系统、网络拓扑的网络空间可视化模型均存在一定的表达缺陷，因此亟需构建网络空间自身的坐标系作为网络空间信息系统的参照基础。其中虚拟复杂的网络空间特性为坐标系设计带来了几个重大挑战。1) 网络空间的空间维度的多少。2) 是否存在类似经纬度的恒定编号系统可作为基础的空间维度向量。3) 如何定义基础二维、三维坐标系。4) 如何表达网络空间的层次性、区域性特点。本节主要围绕上述问题探讨网络空间本源坐标系的构建方法，设计网络空间统一绘制背板。

4.1　基础坐标向量

网络空间坐标系设计的关键在于基础坐标向量的选取，在地理信息系统中体现为经度和纬度。考虑到网络空间的空间特性，基向量的选取需要遵循正交、恒定以及层次化、细粒度揭示网络空间的原则，并思考如何在该向量空间中表征网络异构载体、数据、用户和交互操作等。

根据某高校关于网络空间坐标系和网络空间地图的研究发现[4]，在复杂多变的网络空间中，IP 地址、逻辑端口地址、自治系统编号（ASN, autonomous system number）等一些恒定的编号系统，适合作为网络空间维度向量的候选者。其中，基于 IP 地址映射网络空间不仅支持网络资源要素的定位，而且允许表示坐标系中主机之间的网络通信交互。逻辑端口作为识别网络服务的标识，虽然无法在网络空间中唯一定位要素的空间位置，但是可以在 IP 地址的基础上实现特定需求下的网络应用层信息的表达。AS 作为网络空间域间业务往来和通信的基本单位，将其设计为基础坐标向量有利于描述 IP 地址空间的自治域聚合特征以及 AS 层次的空间信息。

4.2　坐标系设计

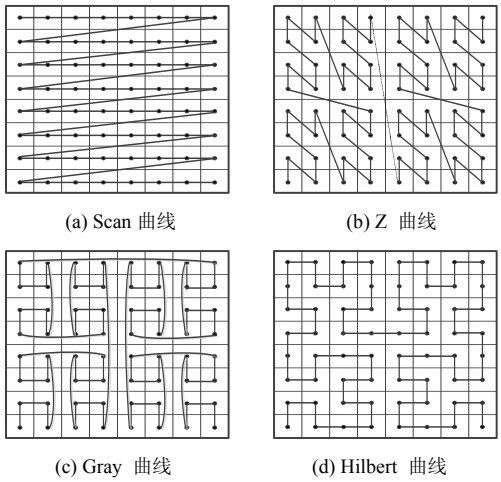
进一步地，在给定坐标向量的基础上，本节将详细描述网络空间二维、三维坐标系的设计，同时构造映射算法使其能够控制坐标系统参数，实现空间信息的层次化表达，并满足坐标平面的连续性和聚合性，以反映网络空间区域的概念。

4.2.1　IP 坐标系

本文借鉴 Irwin 等[18]的研究以及某高校研究[4]中有关 IP 坐标系的设计思想，构建以 IP 地址为基础的二维网络空间坐标系作为网络空间信息系统模型的参照基础。考虑到一维 IP 坐标系以直线和点的形式表达网络空间相对离散且不太直观，因此，需要引入空间填充算法实现低维空间和高维空间之间关联映射，将网络空间聚类到 $2^n \times 2^n$ 的二维 IP 地址空间平面，其中 n 表示二维空间阶数。

如图 1 所示，空间填充曲线主要包括 Scan 曲线、Z 曲线[19]、Gray 曲线[20]和 Hilbert 曲线[21]。最简单的 Scan 曲线将一维空间中值为 S 的点映射到二维空间中，其坐标为 $<x,y>=<S \bmod 2^n, S \operatorname{div} 2^n>$。Gray 曲线基于格雷码，通常采用递归方式进行构造，为获取 m 阶 Gray 曲线，需将一阶 Gray 曲

线网格由(m−1)阶 Gray 曲线进行旋转填充，其一维空间中的值 S 可通过对网格坐标的格雷码进行位交叉操作得到。Z 曲线、Hilbert 曲线对于给定的象限，其绘制方式与 Gray 曲线类似，同样是由象限的位置以及所在大方形的曲线走势所决定。



(a) Scan 曲线　(b) Z 曲线　(c) Gray 曲线　(d) Hilbert 曲线

图 1　空间填充曲线（n=3）

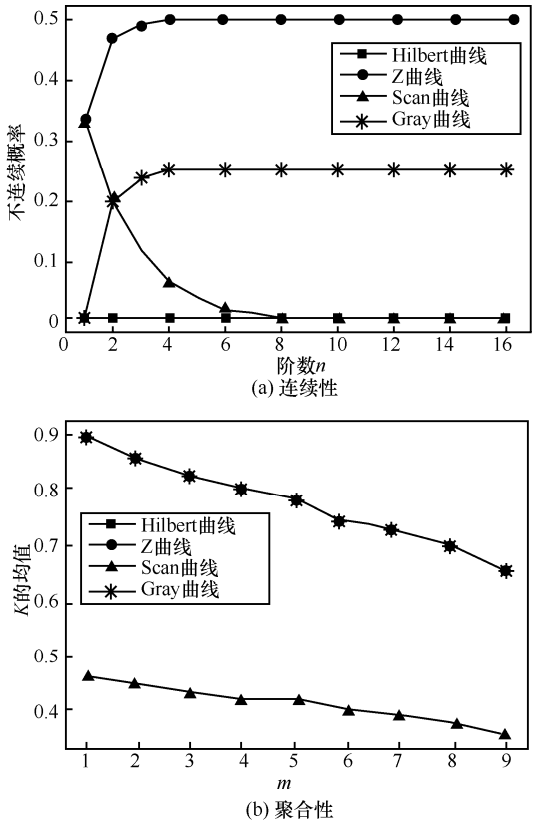
网络空间二维 IP 坐标系映射算法的选取应遵循一定的层次性、连续性以及聚合性原则。具体定义如下。

层次性。网络空间由复杂多样的异质网元构成，需要通过控制填充曲线的阶数 n 以表示网络空间的不同层次粒度。如图 1 所示，n=3，可将 IPv4（Internet protocol version 4）地址空间划分为 64 个子网，其中每个子网的网络部分地址按照填充曲线构造方式排列。观察发现，除 Scan 曲线外，其余映射算法在缩放过程中随着 n 的增加，对每个子网的网格依次进行层次分解（四等分），在保留子网相对位置不变的同时，可得到该 IP 地址前缀下的网络展开。直到 n=16 得到整个 IPv4 地址空间，符合网络空间在不同网络层次、观察粒度下的资源要素的表达需求。

连续性。希望一维连续 IP 地址/网络部分地址在二维空间中也连续，保留曲线上网络空间数据的局部性。

定义 1　一维空间 R 与二维空间 Rd 之间的一一映射记作 Q，若对于 $\mathrm{IP}_1 \in R, \mathrm{IP}_2 \in R, \mathrm{IP}_1 \neq \mathrm{IP}_2$，且 $d(\mathrm{IP}_1, \mathrm{IP}_2) \leqslant 1$，映射后 $d(Q(\mathrm{IP}_1), Q(\mathrm{IP}_2)) \leqslant 1$ 记为 IP_1 和 IP_2 连续，否则记为不连续，其中 d 为欧氏距离。如图 1 所示，Scan 曲线、Z 曲线、Gray 曲线均存在“跳跃”，Hilbert 曲线则处处连续。通过改变阶数 n 计算不连续概率，如图 2(a)所示。从图 2(a)中可看出，Hilbert 曲线的连续性最优，Scan 曲线次之，Z 曲线的不连续概率最高。

聚合性。为了反映网络空间新的区域组织形式，在二维空间中，填充曲线最好能够体现较高程度的区域 IP 地址聚合的特性。具体地，在整个 n=16 的 IPv4 地址空间中，给定随机起点的正方形区域 $D(2^m \times 2^m)$，设 D 内所有 IP 地址聚合得到的最长子网掩码长度为 L，对于给定的 m，L 越大表示聚合粒度越细，且 L 最大为 $32-2m$，引入 $\dfrac{L}{32-2m}$ 作为聚合性评价指标 K。通过改变区域 D 大小，执行蒙特卡洛随机模拟法，得到 K 的均值随 m 变化的曲线如图 2(b)所示。从图 2(b)中可看出，Hilbert 曲线相比于 Scan 曲线的聚合程度更高。



(a) 连续性

(b) 聚合性

图 2　评价指标

综上所述，对比 4 种空间填充曲线，Hilbert 曲线映射算法在表达网络空间的层次性、连续性、聚合性上存在一定的优势。以 IPv4 地址空间中 10.0.0.0/24 子网为例，基于 Hilbert 曲线映射算法构建的二维 IP 坐标系如图 3(a)所示。最细粒度下每个网格代表一个 IP 地址，观察发现子网 10.0.0.0/26、

10.0.0.64/26、10.0.0.128/26、10.0.0.192/26 下的 IP 地址相邻且聚合成网络区域，能够较好地反映网络空间的特性和规律。

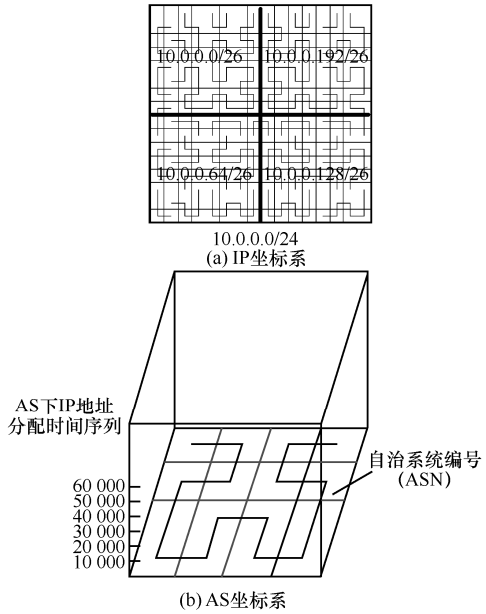


图 3　网络空间坐标系

4.2.2　AS 坐标系

4.2.1 节所述以 IP 地址为基础的二维坐标系可以满足多数场景下网络空间的表达需求，然而受限于部分 AS 的 IP 段分配并不连续，例如，AS4809 下分配 241 个不同的 IPv4 地址段，这些离散的地址段在二维 IP 坐标系中的聚合程度不高，使 IP 坐标系对网络空间 AS 级别流量、拓扑以及攻击特征等信息的表达缺乏一定的直观性。

本文在之前关于网络空间 AS 坐标系研究的基础上[4]，进一步丰富了第三维坐标的含义，用作之后网络空间信息系统的另一参照基础。AS 作为单一管理体系下多个路由器的集合，是网络空间域间业务往来和通信的基本单位。上述空间填充曲线均可作为升维算法将指定范围的一维 ASN 映射到二维 AS 坐标系空间，本文仍选择 Hilbert 曲线构建 AS 坐标系的基础二维平面。此外，考虑到二维 AS 坐标系局限于 AS 要素的表达，而 IP 地址作为网络空间的关键要素难以得到体现，因此可添加 AS 下 IP 地址分配的时间序列作为第三维矢量与之正交构建三维 AS 坐标系，进一步描述网络主机间的信息交互，具体示意如图 3(b)所示。

5　网络空间信息系统

网络空间信息系统与地理信息系统类似，都是基于计算机对空间信息数据进行分析、处理、存储和可视化的工具，其中空间信息定义为具有坐标定位的实体之间的联系以及相互作用的表征。

网络空间信息系统与地理信息系统不同，地理信息系统以地理坐标系为基础，网络空间信息系统以网络空间坐标系为基础。

本节主要研究如何在网络空间自身坐标系的基础上，设计网络空间信息系统模型。

5.1　网络空间信息系统模型

网络空间信息系统的基本设计要点是，在选定的网络空间坐标系基础上，支持对网络空间的多尺度、多维度和多视图表达，并且能够实现网络空间信息系统与地理信息系统之间的映射。

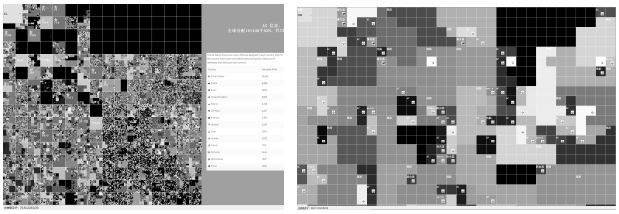
1) 网络空间的多尺度表达设计

多尺度表达是人类由远及近认知空间信息的基础。地理信息系统中，比例尺决定了实地轮廓转变为制图表象的缩小程度，在不同的尺度下地理要素及其属性特征会发生不同程度的聚合或分裂[22]，使地理信息内容的显示呈现一定的层次性。

网络空间信息系统也应具备对海量资源要素在不同粒度下的定位和表达能力。以二维 IP 坐标系为例，其定义模型保证了坐标系的可伸缩和层次化特性，参照地理测度量纲系统，本文初步进行了网络空间信息系统模型的比例尺设计，将网络空间的层次结构划分三层：自治域、网络、主机，实现网络空间全域覆盖和本源描述。

图 4 通过在 IANA（Internet assigned number authority）搜集数据，从多个比例尺尺度对网络空间要素进行描述与表达。首先从自治域层完成对整个网络空间的映射与划分，图 4(a)中不同灰度值表示不同网络空间要素 AS 映射在二维 IP 平面中的区域位置，不同符号代表 AS 名称，右侧显示 AS 的统计数据，其中比例尺 1:/20 下一个像素点实际为包含 2^{12} 个 IP 地址的子网。接着可伸缩表示自治域在网络层次结构下的展开，在 1:/24、1:/28 比例尺下分别采用不同灰度表示 AS 下的骨干网、接入网、物联网等 IP 地址分布信息，进而细粒度表示部分网络层资源要素的区域性，其中地理空间中分散的点在网络空间中可能聚合成块。图 4(b)则基于多尺度认知下的最小观测粒度，实现对网络层的进一步展开，结合网络空间要素组件探测和识别技术[23-24]，重点突出校园网下主机级别的属性聚合特征，如链路、路由器、交换机等，符合人类由远及近认知网络空间

信息的思维，同时可以满足不同管理人员的可视化需求。



(a) 比例尺 1:/20　　(b) 比例尺 64:/32

图 4　网络空间要素定位和表达

2) 网络空间的多维度表达设计

具体设计了 2 种多维度表达模型：一种与地理信息系统类似，通过在基础地图上叠加图层的方式实现多维度信息的表达；另一种与地理信息系统不同，通过添加新的空间坐标维度刻画网络空间的多维度信息，在物理空间中，很难添加三维以上的空间维度，而在网络空间这个虚拟空间却很容易添加新的空间维度，因此基于坐标维度扩展的表达方法是网络空间信息系统的重要特征。本文以网络流量专题地图和网络拓扑专题地图为例，阐述了在 IP 坐标系以及 AS 坐标系底图的基础上，添加第三维坐标系或叠加图层的多维度表达方法。

① 网络流量专题地图

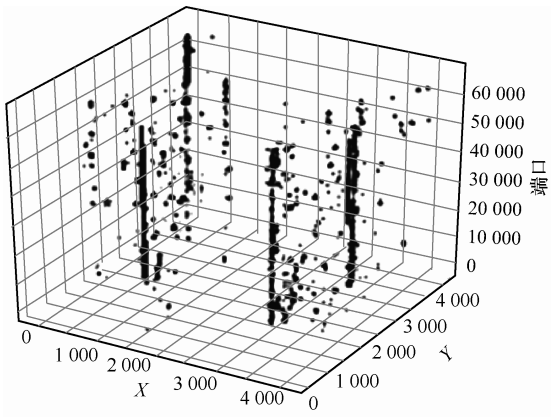
为了反映网络空间中流量的分布规律及其相互关系，本文将识别服务的逻辑端口（port）分别与 IP 地址、ASN 正交构成三维坐标系，设计网络空间流量专题地图。

图 5(a)显示三维 IP-Port 坐标系下访问某主页的源 HTTP（hypertext transfer protocol）流量分布，数据来源于某学校门户网站被动采集的网络日志[25]。以 A 类地址范围 1.0.0.0/8 为例，在比例尺 1:/32 下，（x,y）表示某一 IP：port 一天内发起 HTTP 请求的次数，其在三维坐标中的分布情况反映了该主页的流行程度以及用户请求状态，观察发现源 IP 地址流量主要集中在注册端口以及动态端口，并未存在异常流量。通过将基于 TCP/IP（transmission control protocol/Internet protocol）的网络应用服务产生的流量绘制在网络空间流量专题地图中，有助于直观表达互联网业务及流量往来情况。类似地，在二维 AS 坐标系的基础上添加端口维度与之正交，能够表示 AS 聚合条件下流量分布及端口使用情况。

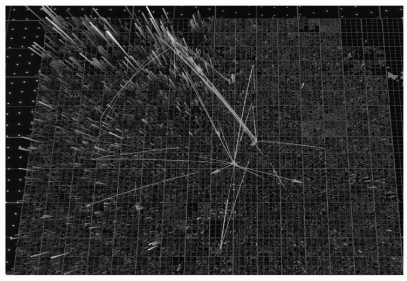
相比于地理信息系统，网络空间流量专题地图摆脱了地理维度的束缚，在网络空间基础底图构建的骨架上，定位和体现与流量专题内容有关的相互关系，协助执行网络监控任务以及与网络安全相关的数据取证。

② 网络空间拓扑专题地图

鉴于网络空间拓扑专题地图需要反映网络拓扑的发展变化以及动态规律，本文采用叠加图层的方式，在保持底图恒定的同时层次化描述实时的拓扑连接信息。图 5(b)以三维 AS 坐标系为例，说明网络空间拓扑专题地图的优越性。具体地，AS 间流量往来关系定义了高级别的全局互联网拓扑，可以帮助用户深入了解网络对等生态系统的技术、安全等需求。图 5(b)在 AS 坐标系上叠加图层，采用飞线表示部分自治域之间的流量往来，发现 IP 数目较多的 AS 往往具有较丰富的拓扑连接关系。与传统 AS 层拓扑模型相比，网络空间拓扑专题地图提供了详细描述和表达的能力，而非使用抽象的点和线来表示拓扑关系。



(a) 网络流量



(b) 网络空间拓扑

图 5　专题地图

此外，通过在 IP 坐标系的基础上叠加拓扑图层，可伸缩实现不同粒度下的网络拓扑表达，包括网络拓扑、路由拓扑以及主机拓扑等，协助不同级别的网络管理人员检查硬件配置情况，及时发现网络中的瓶颈和故障。

3) 网络空间的多视图表达设计

本文重点关注网络空间与地理空间之间的相互映射和协同表达，同时和同步使用网络空间地图和地理地图表达同一对象。观察发现，地理地图和网络空间地图在空间信息表达形式上的侧重点有所不同，图 6 描述了一个/16 子网下的不同任播地址（黑色和灰色）在网络空间的分布情况。任播作为一种网络寻址和路由策略，设计用于将单个客户端路由到“最近”的服务器节点，其中一组服务器节点共享相同的任播地址。很明显，任播地址在地理地图中往往呈现广泛的离散分布状态，而在网络空间地图中收敛至一个点（/24 子网地址），体现了对相同数据集[26]的不同呈现方式。

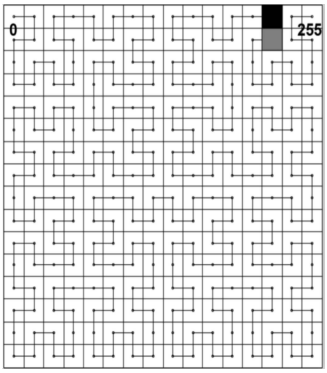


图 6　网络空间地图下任播地址的分布

结合网络空间地图和地理空间地图同时表达网络空间要素提供了一种多视图全面观察网络空间的方法，具体映射可通过对 IP 地址定位库[27-30]执行数据库操作实现。其中难点在于 IP 地址定位库存在一定的缺失和不准确性，借鉴文献[31]提出的一种基于数据一致性（DCR, data consistency rate）分配投票份额的多数据库融合方法，本文在现有数据库的基础上融合多点 Ping、Traceroute 测量结果，提高 IP 地址定位的精度，用于网络空间信息系统多视图的构建。

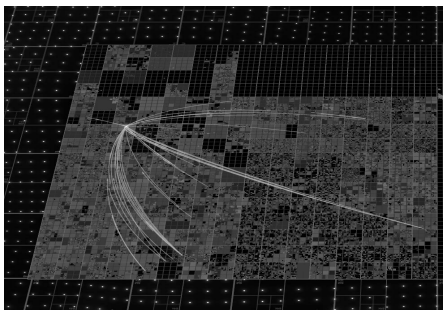
5.2　应用场景实例

以上述坐标系与模型理论为基础研发的网络空间信息系统原型，适用于网络资产管理、网络性能监控、网络安全态势感知等不同场景。本节以 DDoS（distributed denial of service）攻击态势分析和勒索病毒态势分析为例，具体介绍网络空间信息系统原型的应用方式和优势。

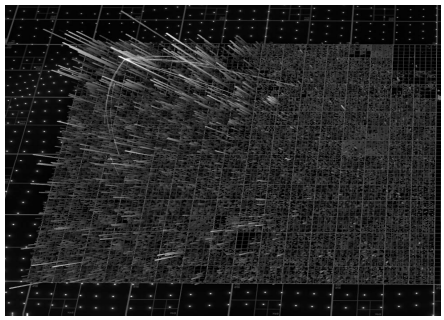
5.2.1　DDoS 攻击态势分析

网络攻击已成为互联网的背景流量，无时不在、无处不在，其中 DDoS 攻击是最难应对的一种安全威胁。

图 7 为网络空间信息系统下，某高校校园主页遭受 DDoS 安全攻击时的可视化表达，其中飞线表示攻击方向，线的粗细表示攻击流量。从图 7(a)中可观察到分散的攻击源 IP 地址段，受益于 IP 坐标系下信息系统模型的多尺度表达，网络安全人员能够基于比例尺伸缩显示傀儡主机所属的自治域、网络、机构、属性类别等网络空间信息，屏蔽攻击源 IP 地址发送的数据分组，从而实现对 DDoS 的有效追踪、防范以及漏洞修复。图 7(b)则对 IP 地址空间进行 AS 聚合，在 AS 坐标系下表达 AS 级别的攻击信息，有助于快速定位攻击源所属的自治系统。进一步在 AS 坐标系下，借助拓扑专题地图显示 DDoS 攻击的 AS 拓扑路径，可实现对攻击源的拓扑溯源和发现，进而指导在攻击时更改互联网基础设施的连接性。



(a) IP 坐标系



(b) AS 坐标系

图 7　DDoS 攻击场景

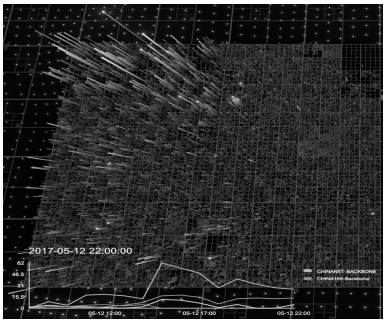
此外，借助多视图表达中地理地图显示的源与目的 IP 的地理分布特性，可以进一步实现基于地理坐标系的定位追踪。

5.2.2　勒索病毒态势分析

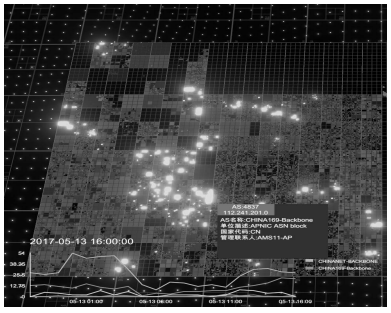
网络空间信息系统可通过添加时间维度，表达网络病毒随时间发展的态势，协助安全管理人员根

据病毒宿主的网络区域的变化建立传播模型，分析病毒内部特征。

2017 年 5 月 12 日，勒索病毒 wannacry 在全球爆发，图 8 绘制了中国范围内 wannacry 感染数据在不同坐标系下的可视化表达，其中左下角显示随时间变化的感染主机数目的统计数据，高亮表示感染的主机 IP 地址段以及对应的 AS。相比于地理信息系统局限于地理信息的表达，网络空间信息系统更能反映病毒感染的时空特性。如图 8(a)所示，wannacry 呈现平稳的增长速度以及持续的感染时间，约 7 h 达到峰值，且不同 AS 均受到不同程度的感染。图 8(b)以 IP 坐标系为背板，观察发现病毒宿主往往分布在相邻的 IP 地址段，表现出一定的网络区域特性，且通过选择表示 AS4837 中国 169 骨干网下对应的 IP 地址分布，可直观呈现该 AS 下感染的主机数量及严重程度。



(a) AS 坐标系



(b) IP 坐标系

图 8 中国范围内 wannacry 感染数据在不同坐标系下的可视化表达

在预防和修复方面，用户可以伸缩显示主机所处的 AS、网络、机构等网络信息，配合网络空间流量专题地图的使用，分析被感染主机的异常流量端口，协助相应的网络管理人员及早防范和发布补丁。

6 结束语

针对网络空间的空间信息系统建模问题，本文分析了传统地理信息系统和网络拓扑信息系统在描述网络空间方面的不足，以网络空间坐标系为基础坐标系，提出了一种多尺度、多维度、多视图的网络空间信息系统模型，满足不同网络层次、不同观察粒度下的网络空间要素的表达需求。本文还进一步阐述了网络空间地图比例尺、网络空间专题地图以及网络空间地图与地理地图的映射等关键技术，并基于上述理论和技术基础，实现了网络空间信息系统的原型系统，应用于 DDoS 攻击态势分析、勒索病毒态势分析等场景。实际效果表明，网络空间信息系统与传统地理信息系统相比，更有助于直接表达网络空间特性，更便于网络空间事务的展现和处理。

参考文献：

[1] MÜLLER-MAGUHN A, POITRAS L, ROSENBACH M, et al. Treasure map: the NSA breach of telekom and other German firms[J]. Feedback, 2014, 12: 13.

[2] LEE N, PLAN X, GENERATION Z. Counterterrorism and cybersecurity[M]. Berlin: Springer, 2015: 301-319.

[3] KSHETRI N. Kaspersky lab: from Russia with anti-virus[J]. Emerald Emerging Markets Case Studies, 2011, 1(3): 1-10.

[4] MIAO C C, WANG J L, ZHUANG S Y, et al. A coordinated view of cyberspace[J]. arXiv Preprint, arXiv:1910.09787, 2019.

[5] JIANG B, ORMELING F J. Cybermap: the map for cyberspace[J]. The Cartographic Journal, 1997, 34(2): 111-116.

[6] 张峥. 赛博地图构建理论研究[D]. 郑州:信息工程大学, 2012.
ZHANG Z. The research on theory of cybermap[D]. Zhengzhou: Information Engineering University, 2012.

[7] GUONIAN L Ü, LIN W Y, ZHAO Y U. Surveying and mapping geographical information from the perspective of geography[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1549-1556.

[8] 郝梁. 电子地图与地理信息系统[J]. 电子技术与软件工程, 2019(11): 249.
HAO L. Electronic map and geographic information system[J]. Electronic Technology & Software Engineering, 2019(11): 249.

[9] BATTY M. Virtual geography[J]. Futures, 1997, 29(4/5): 337-352.

[10] BAKIS H. Understanding the geocyberspace: a major task for geographers and planners in the next decade[J]. Netcom, 2001, 15(1/2): 9-16.

[11] KUIPERS J H. Analyzing the K-root DNS anycast infrastructure[C]//Twente Student Conference on IT. 2015: 1-6.

[12] HUFFAKER B, PLUMMER D, MOORE D, et al. Topology discovery by active probing[C]//2002 Symposium on Applications and the Internet (SAINT) Workshops. IEEE, 2002: 90-96.

[13] FUKUDA K, CHO K, ESAKI H. The impact of residential broadband traffic on Japanese ISP backbones[J]. ACM SIGCOMM Computer Communication Review, 2005, 35(1): 15-22.

[14] SHANNON C, MOORE D. The spread of the witty worm[J]. IEEE Security & Privacy, 2004, 2(4): 46-50.

[15] MAHADEVAN P, KRIOUKOV D, FOMENKOV M, et al. The Inter-

net AS-level topology: three data sources and one definitive metric[J]. ACM SIGCOMM Computer Communication Review, 2006, 36(1): 17-26.

[16] KEYS K, HYUN Y, LUCKIE M, et al. Internet-scale IPv4 alias resolution with MIDAR[J]. IEEE/ACM Transactions on Networking (TON), 2013, 21(2): 383-399.

[17] JIANG J, XU X L, CAO N. Research on improved physical topology discovery based on SNMP[C]//IEEE International Conference on Computational Science and Engineering. IEEE, 2017:219-222.

[18] IRWIN B, PILKINGTON N. High level internet scale traffic visualization using hilbert curve mapping[M]. Berlin: Springer, 2008: 147-158.

[19] ORENSTEIN J A, MERRETT T H. A class of data structures for associative searching[C]//The 3rd ACM SIGACT-SIGMOD Symposium on Principles of Database Systems. ACM, 1984: 181-190.

[20] FALOUTSOS C. Gray codes for partial match and range queries[J]. IEEE Transactions on Software Engineering, 1988, 14(10): 1381-1393.

[21] MOON B, JAGADISH H V, FALOUTSOS C, et al. Analysis of the clustering properties of the hilbert space-filling curve[J]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13(1): 124-141.

[22] 贾奋励. 电子地图多尺度表达的研究与实践[D]. 郑州: 信息工程大学, 2010.
JIA F L. Research and practice of multiscale expression of electronic maps[D]. Zhengzhou: Information Engineering University, 2010.

[23] DURUMERIC Z, WUSTROW E, HALDERMAN J A. ZMap: fast internet-wide scanning and its security applications[C]//USENIX Security Symposium. 2013: 47-53.

[24] ADRIAN D, DURUMERIC Z, SINGH G, et al. Zippier ZMap: Internet-wide scanning at 10 Gbps[C]//Usenix Conference on Offensive Technologies. USENIX Association, 2014: 8.

[25] LIU S C, WANG J L, WANG H, et al. WRT: constructing users' web request trees from HTTP header logs[C]//2019 IEEE International Conference on Communications (ICC). IEEE, 2019: 1-7.

[26] CICALESE D, AUGÉ J, JOUMBLATT D, et al. Characterizing IPv4 anycast adoption and deployment[C]//The 11th ACM Conference on Emerging Networking Experiments and Technologies. ACM, 2015: 16.

[27] LI H, HE Y, XI R, et al. A complete evaluation of the Chinese IP geolocation databases[C]//2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA). IEEE, 2015: 13-17.

[28] POESE I, UHLIG S, KAAFAR M A, et al. IP geolocation databases: unreliable?[J]. ACM SIGCOMM Computer Communication Review, 2011, 41(2): 53-56.

[29] GHARAIBEH M, SHAH A, HUFFAKER B, et al. A look at router geolocation in public and commercial databases[C]//The 2017 Internet Measurement Conference. ACM, 2017: 463-469.

[30] SONG J, XU K, SONG M, et al. Credibility evaluation method of domestic IP address database[J]. Journal of Computer Applications, 2014, 34: 4-6.

[31] LI H, ZHANG P, WANG Z, et al. Changing IP geolocation from arbitrary database query towards multi-databases fusion[C]//2017 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2017: 1150-1157.

[作者简介]

王继龙（1973− ），男，黑龙江大兴安岭人，博士，清华大学教授，主要研究方向为网络空间治理、网络空间测绘、位置网、未来网络试验设施等。

庄姝颖（1996− ），女，河南周口人，清华大学博士生，主要研究方向为网络空间测绘。

缪葱葱（1993− ），男，浙江台州人，清华大学博士生，主要研究方向为网络空间测绘、网络管理、网络数据挖掘等。

安常青（1970− ），女，江苏徐州人，清华大学副研究员，主要研究方向为计算机网络体系结构、网络协议、网络测量。