暗网环境下恐怖主义信息挖掘与分析 *

谢　玲

【内容摘要】 随着互联网和人工智能的快速发展，恐怖主义活动方式变得更为隐秘。网络恐怖主义活动空间从表网的在线社区与视频分享平台转向暗网黑市交易和暗网论坛。针对暗网节点通信和交易信息高度匿名等特点，相应的反恐工作需要采取新的技术手段，以深度切入恐怖分子用于扩大影响的暗网空间。为此，可以设计和优化专门适用于暗网的聚焦爬虫工具，挖掘恐怖主义活动信息；可以搭建涉恐事件数据模型，分析恐怖组织可能的线下攻击手段；可以利用社会网络分析和超链接分析方法，定位暗网内恐怖主义隐蔽社区和危险用户集群。通过准确把握暗网中的匿名恐怖主义活动规律，抓住安全监控与防范的重点、难点，综合施策，以及适时制定和精准实施反恐网络安全策略，能够最大限度遏制暗网恐怖主义的危害，保护国家安全和人民利益。

【关键词】 恐怖主义　暗网　“洋葱”路由器　爬虫工具　社会网络分析

【作者简介】 谢玲，西南政法大学国家安全学院讲师、西南政法大学总体国家安全观研究院助理研究员（重庆　邮编：401120）

【中图分类号】 D815.5　　**【文献标识码】** A

【文章编号】 1006-1568-(2021)03-0135-17

【DOI 编号】 10.13851/j.cnki.gjzw.202103008

* 本文系 2020 年西南政法大学总体国家安全观研究院生物安全风险防控和治理体系建设专项研究课题“城市生物恐怖袭击 AHP 潜在攻击风险模型及防控策略研究”（JS-ZTGJAOG-005）的阶段性成果，同时也是同年 8 月由《国际展望》《国际安全研究》及西南政法大学国家安全学院共同举办的第 8 届国际安全研究论坛的参会论文。

基于互联网访问权限的若干分层与互联网技术的深度应用，全球恐怖主义活动已深入到不同层级的网络空间，形成了活动在表网（Clearnet）中的恐怖主义和活跃于暗网（darknet）中的恐怖主义，网络恐怖主义发生空间转换、大量寄生于暗网渠道。一方面，这是因为强大的反恐执法措施对恐怖主义的表网生存空间产生强烈挤压，恐怖分子面对随时被追踪、定位的风险，必然转而在新的网络生态环境中实施在线恐怖主义活动。另一方面，暗网所具有的通信和交易隐藏特点可以降低恐怖组织及其成员在表网活动时的法律风险。恐怖分子使用匿名通信软件与比特币等加密数字货币支付系统可以逃避表网流量分析，最大限度地减小被发现的可能性。由于这种原因，暗网更适宜于恐怖主义信念、恐怖活动方式等信息的留存和累积，恐怖组织发布宣传材料等活动更无顾忌，筹款、招募、协调的行动更加便利。这些都使得暗网成为恐怖主义犯罪的“天堂”。

正因为如此，观察和分析恐怖主义活动在暗网中的规律及其信息，可以系统准确评判恐怖组织的网络技术成熟度，了解其在暗网扩张的手段和策略，并预测和掌握其线下活动的计划。暗网的加密特性与恐怖主义活动的隐蔽性一经结合，建立针对暗网的恐怖主义信息自动采集机制就会更加困难。本研究拟在这方面进行一些探索，并尝试提出以下综合性思路。一是建立一个初步的暗网反恐信息挖掘与分析系统，从信息爬取、事件预测和网络交互三个角度研究恐怖组织的暗网使用情况。二是采取包括设计暗网聚焦爬虫工具、建立涉恐事件数据模型、分析危险社群用户在内的技术手段，从暗网黑市和暗网论坛中收集、提取恐怖主义活动信息，分析和预测恐怖组织的在线状况、发展动态与线下行为。这样的暗网信息工作有助于实务部门研判暗网恐怖主义活动规律，制定更加全面、有效和专业的反恐行动对策。

一、网络恐怖主义的滋生与蔓延

随着互联网通信和人工智能技术的迅速发展，年轻人群在线上的活动、社交方式以及接收信息的途径发生显著变化，恐怖组织开始以网络和大数据

为中心，有系统地使用互联网输出恐怖主义信息，利用多层网络结构发展出各种灵活的线上组织形式。例如，建立恐怖组织“官方”网站、创建社交媒体账号发布随时可用的在线宣传资源，以影响低龄激进分子思想，吸引更多的支持者。互联网逐渐演变为融资、招募、武器交易等恐怖主义活动的常规线上支持工具，恐怖主义更有向互联网空间的隐秘角落蔓延的趋势。其中，网络深层的暗网站点成为恐怖主义最新的社交招募和信息输出路径。在互联网空间中，恐怖组织逐步构建起一个由恐怖组织网站、分散的网络社交媒体和暗网三部分构成的庞大的实时在线恐怖主义信息网络系统。为行文方便，本文将暗网环境下的恐怖主义简称为“暗网恐怖主义”。

21 世纪初，有国外学者指出，恐怖组织试图开发一个在全球范围内指挥和控制成员的网络通信情报系统，包括但不限于建立专门的网站作为恐怖主义信息发布平台。[①] 基于此，最早开始关注恐怖主义网络活动的一批美国学者依据本国安全机构提供的恐怖组织名单，登录部分恐怖组织设立的“官方”网站进行信息采集和研究。他们利用内容分析技术对网站内容进行动态追踪和多维研判，试图了解恐怖分子如何以互联网为平台进行恐怖主义活动的宣传、策动与实施。[②] 该项研究对于分析传统恐怖组织的网上活动规律具有重要价值，但其局限性在于通过建立“官方”网站的形式实施在线宣传的恐怖组织，不到名单显示的恐怖组织数量的一半，对未建立专门网络平台发布信息的恐怖组织，则无法借此观察其线上活动的新特点和新趋势。

随着网络社交媒体的快速发展，恐怖分子已经大量借助脸书（Facebook）、照片墙（Instagram）、推特（Twitter）、优兔（YouTube）等社交媒体账户和标签、网络论坛、游戏模块宣传恐怖主义，形成了恐怖主义线上活动的新趋势。据统计，极端恐怖组织“伊斯兰国”（Islamic State of Iraq and Syria）在全球拥有近 1 000 家社交和数字媒体运营商，创办出版了在精致程度上可与主流期刊比肩的英文杂志《达比克》（*Dabiq*），从事相

① Hsinchun Chen et al., *Terrorism informatics: Knowledge Management and Data Mining for Homeland Security*, Berlin: Springer Publishing Company, 2005, p. 2.

② Gabriel Weimann and Yariv Tsfati, “Terrorism.com: terror on the Internet,” *Studies in Conflict and Terrorism*, 2002, pp. 317-332.

关工作的人员数量超过许多大型公关机构。[①] 此外，恐怖组织还在各个访问流量较高的社交媒体上注册账号、发布恐怖主义信息，进行宣传、筹款、协调、招募人员等活动，扩大社会影响。以 YouTube 视频为例，截至 2021 年，每月有 23 亿用户观看总时长达 10 亿小时的视频，占全球互联网访问量的一半，其中 18 至 25 岁的互联网用户是 YouTube 的使用主力。[②] 由于 YouTube 用户量大、视频发布门槛低、不注册账号即可访问，为恐怖组织上传非法暴力视频、分享极端主义内容、连接其他煽动宗教和民族仇恨的团体共同形成危险网络社区提供了便利，也为非法活动走向线下提供了支持。据相关研究预测，如果恐怖分子利用 YouTube、Twitter 等社交媒体传递信息，组织、煽动恐怖主义行为，其响应者可在两小时内引发骚乱，对于密切联络的激进网上社群，骚乱的发生时间能缩短到 20 分钟左右。[③]

由于 YouTube、Twitter 每分钟可上传 100 小时视频并能随时删除发布内容，研究者使用常规的分析方法已跟不上海量数据的更新和销毁速度，更难预测相关线下恐怖活动走向。一些学者开始采用爬虫技术检索、发现 YouTube 用户文件中的恐怖主义信息，[④] 结合社会网络分析方法 [⑤]（Social Network Analysis），以节点交互的结构图映射社交网络中恐怖分子及其支持者，分析极端社区的发起形式与组织架构。但是，随着恐怖主义线上活动逐渐由表网向暗网迁移，恐怖分子开始使用隐蔽在线应用程序。前述研究以及以此为基础的反恐安全监控与网络执法打击已无法掌控恐怖组织输出暴力极端主义的完整路径，即使采取常规方式扫描、发现和分析恐怖主义触角所伸向的暗网空间也会存在遗漏。只有利用技术手段进一步深入恐怖分子倚重的暗网空间，才能对恐怖主义活动进行更系统和有效的针对性打击。

① Oz Sultan, “Combatting the Rise of ISIS 2.0 and Terrorism 3.0,” *Cyber Defense Review*, Vol.2, 2017, p. 44.

② Statista, “Number of YouTube Viewers in the United States from 2018 to 2022 in millions,” https://www.statista.com/statistics/469152/number-youtube-viewers-united-states/.

③ Oz Sultan, “Combatting the Rise of ISIS 2.0 and Terrorism 3.0,” p. 45.

④ Swati Agarwal and Ashish Sureka, Topic-Specific YouTube Crawling to Detect Online Radicalization, 10th International Workshop on Databases in Networked Information Systems, 2015, pp. 133-149.

⑤ Aparna Basu, “Social Network Analysis: A Methodology for Studying Terrorism,” *Social Networking*, 2014, p. 215.

二、暗网结构与暗网恐怖主义的关联

暗网恐怖主义活动已经构成一种新的战略性安全威胁。对于这一隐秘的网络空间，我们要首先查明暗网能够为恐怖组织提供何种支持，查清高风险涉恐人员对于暗网的可及性和通联意图，查实恐怖主义在暗网上的运作模式。只有充分把握暗网恐怖主义的外部行为特征与内部作用机制，才能使信息技术安全策略有针对性地预防和减小暗网恐怖活动可能造成的危害后果。

（一）恐怖主义信息网络与暗网

网络的三层结构中，处于第一层级的表网是指利用标准搜索引擎和网络浏览器可以访问的页面和内容。过去对于恐怖主义与互联网关系的研究主要集中于互联网的表网。处于第二层级的深网（Deepnet）是指托管在互联网开放部分，但却有内容访问限制而未被搜索引擎索引和收录的网络站点，如公司内部使用的网站、网上银行个人账户和图书馆目录等。[①] 但深网仍然可以通过常规网络浏览器和授权的连接方法获得访问。处于第三层级的暗网是隐藏在深网中的一部分特殊的加密子集，它是指由加密网络和匿名通信构建，需要通过“洋葱”路由器（The Onion Router, TOR）、I2P 路由器（Invisible Internet Project）和“随意网”（Freenet）等通信软件提供匿名通信协议才能访问的网络。

以匿名通信软件 TOR 为例，它广泛应用于隐秘的地下自由通信，其“匿名性”的基本原理是通过隐藏流量分析保护用户隐私、支持匿名浏览以及避开网络监控。TOR 网络由成千上万个作为中继节点的服务器组成，每个中继节点都由全球志愿者免费提供。[②] 网络流量经过 TOR 网络从中继节点自动选配的入口节点、中间节点、出口节点时，每一节点都会参与数据包传递的加密，既不记录流量的来源，也不记录流向，IP 地址只显示退出节点，从

① John Robertson, et al., “Darknet Mining and Game Theory for Enhanced Cyber Threat Intelligence,” The Cyber Defense Review, 2016, p. 95.

② 祝世雄：《网络攻击追踪溯源》，国防工业出版社 2016 年版，第 288 页。

而达到隐藏信息发送用户真实 IP 地址的目的。[①] 网络匿名协议覆盖暗网环境下与恐怖活动有关的地下交易市场、在线社区，能够为地理位置分散的恐怖分子及其追随者提供通联渠道与身份保护。加之各国政府过滤和取缔表网中的极端主义内容，频繁变化 IP 地址、令执法追踪困难的暗网成为地下犯罪和恐怖主义活动的“温床”。杰米·巴特利特（Jamie Bartlett）将其描述为“一个能够搜罗网络中所有令人震惊、不安和充满争议的黑暗角落的术语，一个容纳了你所能想象的各种形态的罪犯和捕食者的领地”[②]。

（二）暗网恐怖主义的行为特征

针对不特定多数人施加暴力或以暴力相威胁是恐怖主义的基本行为特征。恐怖组织通过煽动民众、制造恐慌、谋取暴利，以达到某种政治、经济、宗教或社会目的。随着互联网的普及，线下恐怖主义也向线上转移。恐怖组织利用不受追踪的暗网作为恐怖主义信息输出路径，招募恐怖分子、筹措资金，对潜在危险用户施以极端化影响，在条件具备时提供培训和必要帮助，或调集新的恐怖分子参与，对实施线下暴力恐怖主义活动提供便利。基于恐怖组织的暗网线上活动与发生线下暴力恐怖行为相关联的属性，对于暗网恐怖主义目前学界重点关注两个问题。[③]

第一，暗网恐怖主义活动的认定问题。各国学者对于恐怖主义尚未形成统一定义。基于宗教动机、政治意识形态、民族分离主义等复杂动机实施的暴力行为，一般被视为主要的恐怖主义活动类型，但在进行实际认定时，各国对于一些组织的恐怖主义属性问题存在不同意见。例如，美国安全机构认定的国际恐怖主义团体包括黎巴嫩真主党（Hezbollah）、哈马斯卡萨姆旅（Ezzedine al-Qassam Brigades）等在其他一些国家并未被认定为恐怖组织。另有部分学者认为，对于狭义的单一问题，如堕胎、环保、反对执法等引起的极端行径，因存在暴力行为特征以及作为行为后果的恐怖图景，也应属于恐怖主义活动范畴。本文认为，对于这类因单一问题引发的极端行为，应当

① R. Dingledine, N. Mathewson, and P. Syverson, TOR: The Second-generation Onion Router, Proceedings of the 13th Conference on USENIX Security Symposium, 2004, p. 21.

② Jamie Bartlett, *The Dark Net*, London: William Heinemann, 2014, p. 2.

③ 对于极端主义者入侵网络系统、破坏网络设施、盗取重要信息的黑客恶意攻击类网络活动，因不具有恐怖主义暴力行为特征，不在本文研究范围之内。

依据暴力实施者所侵害的具体法益认定为某种类型的普通刑事犯罪，该行为人系刑事法所规制的一般对象；因复杂动机实施暴力的行为人或组织，在聚焦其暴力特性的同时，应以中国国内法为主，结合国际条约和国际社会普遍认可、接受的认定规则判定其组织及活动的恐怖主义性质。认定暗网中的“恐怖主义”不能放大到一个松散而宽泛的范围。

第二，暗网恐怖组织追随者的识别原则。由于受网络恐怖主义洗脑影响的深度不同，部分极端主义分子在现实中发动侵害和袭击的个人动机既多样又相互重叠。据全球应对恐怖主义开源数据库（Global Terrorism Database, GTD）发布的 1970—2016 年恐怖主义意识形态动机背景报告，“意识形态动机未知”类型的恐怖袭击占到所有恐怖主义攻击数量的 24%。[①] 实施此类袭击的危险人员的具体行动指向表面上与恐怖组织并无关联，但通过学习和模仿恐怖分子的暴力恐怖行为方式，其行为的实际表现和危害结果又与恐怖主义活动趋于一致，对其属于恐怖组织成员还是一般追随者在实践中难以准确认定。本文认为，通过暗网站点接触过恐怖组织的用户，只要其主观上表示出认同并自愿接受其意识形态、行为模式和滥杀意图，客观上将暴力行径嵌入各种危险动机，即使仅为了表达其极端想法而实施暴力恐怖行为，一般也应认定为恐怖组织成员并纳入暗网信息监控。对于一些与恐怖组织的隶属关系并不明显，为实施恐怖主义活动而意欲获取相关知识和技能，或者存在购买武器装备意向、可能会实际进行暴力活动的危险个人，则应注意根据情况区分是否纳入暗网恐怖组织追随者的范围。

（三）暗网恐怖主义活动的平台与功能

暗网环境下恐怖分子具体实施的恐怖主义活动类型由暗网平台的构成与功能所决定。暗网平台主要由暗网论坛和暗网黑市两个部分组成，加密聊天室或点对点的即时加密通讯因缺乏网络信息传播的公共特性不列入本文的讨论范围。

第一，暗网论坛。暗网论坛是指利用 TOR 等特殊网络建立的网上信息发布和交流平台。其结构和组织形式与表网论坛类似，由不同主题的分区版

① Charlinda Santifort,Todd Sandler, and Patrick T Brandt, “Terrorist Attack and Target Diversity: Changepoints and Their Drivers,” *Journal of Peace Research*, Vol. 50, 2013, p. 78.

块和若干子版块组成。但暗网论坛采取不同种类和等级的信息安全机制保护用户隐私安全，大量使用图灵测试防止计算机自动访问。因论坛讨论的话题不能被普通搜索引擎检索，暗网论坛中遍布恐怖主义、武器、毒品、“网络黑灰产”[①] 犯罪等非法内容，且论坛之间能够以超链接方式相互关联。

恐怖组织利用暗网论坛收集信息、招募人员、建设虚拟社区、进行恐怖融资，可绕过在表网上从事同样活动面临的法律风险及执法打击。表网上发布的恐怖主义信息因受举报和执法控制存活期短，往往在恐怖主义“官方”站点投入使用、与支持者刚建立初步连接时就被强制关停，这迫使恐怖组织不得不转向影响较大的网站或社交应用离散化地推送信息。而在暗网空间建立较为固定的社群网络发布恐怖主义信息则具有明显优势。例如，暗网论坛可以设置进入权限和匿名访问。暗网论坛认可的特定用户被自由放行的同时，经过匿名通信处理，可以形成较为固定的、经常交流恐怖主义信息和讨论线下活动的激进网络社群。恐怖组织伺机在暗网中发展狂热的追随者，在线传播简易爆炸物、生化危险制剂制作和使用的培训视频，出售枪械武器及使用手册，推出关于核武器的学习教程，这些在暗网论坛上接受恐怖主义思想灌输和武器培训的追随者相对于表网上的支持者更具有行为攻击性和现实危害性。

第二，暗网黑市。暗网黑市是指利用 TOR 等特殊网络搭建的非法商品网络交易平台，其网络销售模式与表网购物平台类似，提供数字货币电子钱包托管服务、国际信用卡资金兑换数字货币的汇兑渠道以及匿名购买选项。常见的非法商品主要是毒品、武器、色情和恶意软件产品、公民个人信息等。在暗网黑市贩卖的各类非法物品中，武器是与恐怖主义活动相关性最强、危险性最高的交易对象，其交易成功意味着恐怖组织和个人可能在线下动用武力实现恐怖主义目标。越是为国际社会严格禁止持有、极度危险的高级别威

① “网络黑灰产”系专有名词。根据百度时代网络技术有限公司与公安部第三研究所网络安全法律研究中心编制发布的《2020 年网络黑灰产犯罪研究报告》，“网络黑灰产”定义为借助互联网技术、网络媒介，为黑客攻击、网络黄赌、网络诈骗、网络盗窃、网络水军等违法犯罪活动提供帮助，并从中非法牟利的犯罪产业。在暗网上贩卖黑客攻击软件、公民个人信息的不法行为属于网络黑灰产。

胁性武器，越有可能为了避开执法视线在暗网空间完成交易流转。[①] 暗网黑市与论坛也有着密切联系，卖家会在暗网论坛零星发布商品信息并提供交易链接，买家循着论坛线程能够查找到暗网黑市买家信息。

通过分析暗网黑市买卖双方购买需求与意向，探知交易指向的高风险地区，再结合其他信息进一步定位其背后存在的恐怖主义组织或极端主义团体，分析该组织运用武器的威胁能级，如是否具有专业技术操作人员、校准维护人员、监测设施和专门场地，可以评估恐怖组织的危险活动等级以及预测恐怖主义行动的发生概率。过去的恐怖组织成员多是一些暴力、无知、拒绝现代化和高科技的宗教狂热分子，但是“9·11”恐怖袭击事件之后美国媒体报道了“一些令人震惊的发现”，恐怖组织招募的年轻一代部分来自中产阶级家庭，他们受过良好的高等教育，能够执行更为复杂的军事行动并使用威胁性武器，[②] 一些恐怖分子甚至认为核武器是他们未来行动的重要手段。因此，在暗网黑市中监测和识别与恐怖主义有关的大规模杀伤性武器交易内容以及核交易企图极为必要。

（四）对暗网信息进行挖掘分析的意义

在 21 世纪初，国内外已有一些关于表网恐怖主义信息挖掘和建模分析的研究，但针对暗网空间的论坛和黑市探讨恐怖主义活动轨迹、交流内容和极端社群结构的研究仍然有限。相对于表网恐怖主义的识别、追踪和预测活动，暗网领域的恐怖主义信息更值得予以关注和深度挖掘。

第一，弥补网络反恐的盲区。在整个实时在线的恐怖主义信息网络系统中，暗网环境下的匿名恐怖主义活动和隐匿踪迹的涉恐交易是反恐监控和安全防范的难点。网络恐怖主义对全球安全的威胁能否消除取决于对网络安全机制最为薄弱环节的把控，因此必须加强对暗网领域恐怖主义繁衍的遏制。

第二，提高信息筛查的精准性。针对表网的恐怖主义信息挖掘一般是以开源出版物、报告、网站和社交媒体为基础进行关键词和联想词的自动或半

① 何由：《隐蔽战争》，中国市场出版社 2013 年版，第 229 页。

② [美]约翰·L. 埃斯波西托、达丽亚·莫格海德：《谁代表伊斯兰讲话》，晏琼英等译，中国社会科学出版社 2010 年版，第 102 页。

自动检索，从海量数据中采集到的大量信息、数据可能都是合法内容和链接。例如，设定检索参数为“与恐怖主义相关的大规模杀伤性武器”，得到的大部分是一般性的知识科普类介绍。而对于社交媒体评论区上传的恐怖视频及讨论发言，爬虫程序很难突破人类语言表达的丰富性、并直接检测评论用户对恐怖主义视频宣传的情感态度，在这两种情形下需要采取手动信息筛查和复杂的文本情绪分析，否则所有相关的站点内容、评论信息都会被认定为符合关键词和联想词特征而纳入数据采集范围，会导致庞大的无效信息大量占据后台解析空间。而暗网中至少有一半登陆者发布非法信息和交易供求任务，[①] 如果统计模型直接从暗网黑市和论坛的线程、文档中构建，涉恐内容出现在筛选集合中的精准性更高。

第三，建立恐怖主义活动的反向推演。我们知道，分析暗网黑市上正在出售的恶意软件产品的种类、特点、入侵途径，涉及机构和个人可以获得网络安全防范的反向提示并以此弥补目标计算机的安全漏洞，以防遭遇现实网络攻击。而对恐怖组织和个人在暗网黑市和论坛上发布的行动指令和交易需求的分析，也可以作为揭示恐怖组织的下一步行动、预测恐怖主义活动的紧急属性和威胁级别的依据，进而有助于相关机构结合其他渠道的调查信息识别具体危险并采取相应干预措施。

三、遏制暗网恐怖主义的技术方略

本研究提出一个与表网信息挖掘和分析相对的暗网技术框架，用以发现、监控、利用暗网环境中的恐怖主义信息，为打击线上、线下恐怖主义寻找突破口。主要技术方略可分为三个部分：一是利用爬虫工具和内容解析器等底层技术设施从暗网黑市和论坛中收集、挖掘、分析恐怖主义信息；二是通过搭建博弈论框架、优化算法分析来暗网数据；三是通过收集的数据发现黑市买家和卖家、恐怖主义信息传播者构成的恐怖主义社交网络。这样，基于实时数据所表征的恐怖活动潜在威胁以及所把握的危险社区和隐蔽恐怖

① Andy Cohen, “Decision-Making Dynamics When Moving Out of Your Comfort Zone,” *Cyber Defense Review*, 2017, p. 54.

社群状况，可以为相关机构提供一个用于预测未来可能发生的恐怖主义场景的数据分析框架，形成打击暗网恐怖主义线下活动的技术支撑条件。

（一）暗网信息的自动抓取与深度挖掘

从暗网空间自动收集和提取恐怖主义信息，建立一个实时变动的暗网数据集合，是进行分析、研究并生成可视化关系图谱的前提。如前所述，一些研究人员在表网上运用文本分析方法和链接分析算法对恐怖组织“官方”网站发布的数据进行收集、存档和研判，并对恐怖分子使用大众社交软件发送离散信息予以密切关注。对于暗网中的黑市和论坛，我们也需要开发专门的爬虫工具来访问涉恐站点、自动获取站点中的敏感信息。

升级版的爬虫系统将抓取站点选择为恐怖主义信息较为集中的暗网论坛、黑市站点，以及在这些论坛网站中收集到的关联站点链接。在设计时应充分考虑暗网与表网信息收集分析系统存在的差异，对暗网爬虫系统的设计进行相应的技术扩容。一是要考虑暗网站点 IP 地址因使用匿名协议变动快，运行稳定性差、使用寿命短，且一般需要使用特定的辅助客户端程序进行访问的特点；二是要在数据抓取时识别和避开暗网上的国际执法钓鱼网站网页；三是要增加相对精准的跨语言检索功能，这是因为暗网论坛除了使用英语进行交流，基于恐怖主义发生的民族和地域性特点，也会使用阿拉伯语、法语、俄语、德语等其他语种发布信息。

第一，暗网爬虫系统主要由爬虫模块和数据仓库模块两大部分构成。爬虫模块包括连接器、下载器以及解析器等组件，负责将目标站点数据下载到本地数据库中。数据仓库模块主要是用于提供各种分析维度的查询接口，并以用户、团体、站点为对象建立模型，通过数据分析评估恐怖主义事件的行为模式、关联模式以及危害程度等，数据仓库模块分析所需要的数据可以定时从爬虫数据库抽取。在两大功能模块之间建立一个“设置”模块，将数据仓库模块初步分析所得的预测结果返回给爬虫模块中的调度器，进一步指引爬虫工具抓取和过滤信息的方向，以提高数据“捕获”的准确性和爬虫系统运行效率。此外，考虑到使用站点的恐怖组织成员分布在世界各地，可能使用多种语言，网络用语与日常沟通用语之间也存在较大区别，使用日常用语

模式进行分析可能会漏掉一些关键信息的收集，因此，可使用接口安装第三方翻译平台的插件来提高爬虫工具在各种语言环境中的应对能力，并通过在系统中自建附加功能模块“语料库”的方式，使系统更好地识别多语种交流中的敏感信息。

第二，爬虫模块的主要构成与功能实现。一是爬虫工具和解析器代码的分离与协调。由于每个站点都需要设计对应的爬虫工具和解析器，为了对不同结构的网站进行数据抓取，需要在设计中分离爬虫程序和解析器的代码，以便日后扩展抓取范围。二者可以通过网络地址管理器（Uniform Resource Locator Manager, URL）和爬虫调度器进行工作协调。二是着重进行 HTML 静态文档与动态数据检索。为解决不同暗网平台的网络结构差异与访问权限控制问题，针对不同站点的特征，使用能够避开访问控制、无响应服务器、去重以及伪装登陆的爬虫工具，从暗网黑市和论坛下载信息。三是恐怖主义活动相关内容的提取。基于每个暗网站点的结构，使用与之对应的内容解析器对下载的 HTML 文档进行解析，将网页内容保存至数据库中，根据有关材料初步判定当前抓取的页面是黑市站点还是论坛网站。其中，暗网黑市与武器交易相关的商品信息、供应商信息、询价信息、论坛宣扬恐怖主义的帖子、作者信息（包括声誉等级、话题兴趣等）、参与讨论的用户信息、超链接等重要字段一经匹配，自动纳入爬虫模块数据库收集储存范围，为后续分析提供数据支持。四是分阶段、分类型下载与分析信息。从爬虫模块细化出能够应对不同类型站点的下载器，刚开始抓取数据时只设置暗网论坛或黑市页面的爬虫工具，对在网站中抓取到的链接，作为链接类型的数据保存在数据库中，以便开发出对应的爬虫工具再提取该链接的网页信息。数据仓库模块应定期抽取数据进行分析，并在该模块建立暗网黑市、论坛用户及团体的数据集市，[①] 以满足暗网恐怖主义社群网络分析的需求。随着系统抓取数据量的增长，可能会发现恐怖分子用于招募人员或购买物资的站点，通过前面数据仓库模块的分析，可以筛选出一批可疑站点继续进行抓取。

① 数据集市是数据仓库的一个子集，是面向某一特定主题的数据中心。为便于分析，整个暗网数据可拆分成“黑市用户”“论坛用户”及“团体”等更小单位，最后再合并为一个能够描述整个事件的宽表。

第三，数据仓库模块的主要构成与功能实现路径。一是数据仓库模块的结构化创建。数据仓库模块的数据来自本系统的爬虫模块，它很少会出现数据结构不一致的情况，但刚开始爬取一个站点时，由于不能准确分辨是否为系统所需要的数据，而假设暗网论坛和黑市中的信息都是比较集中的恐怖主义信息，爬虫工具会将所有发言信息都下载到数据库中。为排除数据源中的非恐怖主义信息，可将数据仓库模块建立为 ODS 数据运营层—DWD 数据明细层—DWS 数据服务层三层结构。由于 ODS 数据运营层与爬虫模块的数据库基本保持相同的数据粒度，在该层不需要进行数据清理，只在数据库表中标记每条数据的关键词。二是 DWD 数据明细层的预处理。DWD 层根据 ODS 层的标记有选择性地保留数据，在数据进入该层之前依据数据仓库技术（Extract Transform Load, ETL）流程进行数据清洗，转换为便于分析的形式。对于比较复杂的信息文本，需要通过特定技术提取关键内容。比如，对于暗网论坛中语义复杂的内容，还需进行关键信息识别，并作为常规的分析数据保存在 ODS 层数据库中；ODS 层再将数据整理成统一的格式以便进行后续分析。比如，一条特定信息除了从其自身内容剥离出可供分析的有效数据外，还与一系列特定信息相关联，这些信息保存在爬虫数据库不同的表结构中。ODS 层根据数据对象之间的关联关系加以延展，可以生成更多用于数据挖掘的重要信息。三是分类过滤无关数据。鉴于爬虫工具获取的暗网论坛和黑市信息可能存在少量与恐怖主义无关的内容——如钓鱼执法网站的信息，可运行分类器过滤功能，并在数据仓库模块中构造分类模型对数据进行信息自动化检测和筛查，[①] 通过自动配置爬虫模块中的调度器，停止与恐怖主义无关网页的数据抓取。四是对 DWS 数据服务层数据进行多维度分析评估。数据仓库模块挖掘的数据源主要是用户发言中涉及恐怖主义招募、活动组织策划方案与武器交易的内容。使用 DWD 层中经过预处理的数据，在关联网页的基础上整理出包含恐怖主义团体、事件类型、地区、策划行动日期、策划时长、武器基本信息等维度的恐怖主义活动事件构成要素，建立多维度可视

① Eric. Nunes et al., *Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence*, IEEE Conference on Intelligence and Security Informatics, 2016, p. 106.

分析的“数据立方体”，[①] 通过“数据立方体”的上卷和下钻，在各个维度对团体危害程度等进行评估。在数据挖掘早期，由于数据量不足且不易评估，主要考虑使用聚类分析算法对人员角色进行基本分类，待后期从其他路径搜集到的信息可以与数据库中爬取的信息进行匹配后，再进一步评估、改进模型，对一连串相关的恐怖主义行为指向进行关联分析，进而使用逻辑回归、决策树和支持向量机等算法对未来的涉恐事件发展趋势进行预测。

（二）暗网信息的智能推演与技术博弈

暗网环境下的信息安全分析，以未知信息价值高于已知事件为假设性前提。因为只有在暗网信息收集、价值排序和关联化处理完成，并形成信息研判结论及转化为相关的行动方案和防御性措施之后，针对暗网信息的分析工作才能对反恐活动发挥真正的辅助作用。即使是涉恐线索暂时缺乏确实证据的支撑，也要尽量从信息分析中得出暴力恐怖事件发生概率和过程的模拟结果。这是因为爬虫工具所抽取信息的规则本身，就是基于暗网中涉恐事件发生的基本框架，只是在信息采集的层面上不能直接显示出明确的恐怖组织行动规则、完整的恐怖主义事件发生逻辑与具体恐怖组织成员和追随者特征。这就需要运用信息分析的假设论和博弈论更加详细地探讨、分析暗网恐怖主义活动可能存在的线下事件模型，并发挥暗网涉恐信息整合的执行意义。

第一，抽取相关性信息构建涉恐事件数据模型。首先，使用爬虫工具遍览暗网中相关链接和网页，提取含有恐怖主义关键词和联想词的各类文档、图片、音频和视频。其次，基于对文本内容类别的筛选，自动识别疑似恐怖分子发布的宣传、招募、培训、沟通、交易等活动信息，以及危险用户的反馈信息。再次，通过数据分析 ODS 层的预处理，将单个零碎的信息数据通过时间、空间和事件等重点分类要素串联为信息链，揭示恐怖组织利用暗网社群建立不同强度的作战组织结构或煽动暴力恐怖活动的可能模式。

第二，探知涉恐事件数据模型发生的有效性。按照博弈论框架下的事件模型及运行结果，预判暗网论坛和黑市恐怖主义在线活动所涵盖的危险要素

① 数据立方体是一种三维数据建模形式，用于数据分析和显示，相当于将一个要素表格以立方体的形式呈现，可视化更强。

及其可能生成的线下攻击行为、规模及涉恐高危人员。以黑市交易为例，一是依据暗网黑市中某种武器或爆炸物的销售量、库存剩余量、数字货币兑换情况，可以估算单项恐怖主义活动成本效益；二是在合理预算条件下，评估恐怖分子购买的武器类型、数量及可能支持何种规模的攻击策略以获取最大的行动回报；三是从爬虫数据库中提取另一组显示主题、位置、日期的强关联性向量增加事件精确度，计算攻击模式下的模拟事件结果以及这一恐怖主义行动的发生概率；四是通过对暗网黑市中交易用户的语言特征、所在地区、交易方式、商品储备类型等信息的标记化和一致性分析，结合数据挖掘判断多个暗网平台上销售的同类型武器商品是否出自同一卖家，确定是否将其识别为高危涉恐成员并进行更为深入的社会网络分析和恐怖分子画像。

第三，推演涉恐事件数据模型与反恐方案的博弈结果。基于对立双方相互作用的博弈论框架，结合爬虫数据库建立数学模型，对包括人员、组织、能力、威胁等级、防御级别在内的恐怖主义行动选择与反恐方案进行赋值评估，以便研究恐怖组织行动的概率、类别和对策。[①] 首先，将爬虫工具抽取到的分类恐怖信息及可能暗含的攻击方式，按照事件关联性逐项建模作为给定变量，确定权重。其次，围绕恐怖组织实施暴恐行动的事件模型，对武器购置、人员招募、信息交易、恐怖融资和隐秘社区的建设维护等事件要素赋值，计算事件模型所代表的整体危险水平。再次，结合其他信息来源的危险变量及赋值、地理信息的实时来源反复调整反恐预案，设置阻截要素以降低恐怖组织可能实施的暴恐活动的发生。最后，对博弈论框架中的双方活动进程进行预案性能评估分析，以及得出调整后的应对措施能否有效控制危险事件的结论。

（三）暗网恐怖主义社群的用户分析

执行爬虫信息抓取分析以及反恐智能推演，主要是针对暗网恐怖组织可能实施的线下活动进行预测性的危机描述、事件识别与预警处置。由暗网的结构特性出发，进一步扩展出这些潜在恐怖主义活动的运作方式和发生顺

① 尹秋菊：《基于信号博弈和 MAS 的交易行为研究》，北京理工大学出版社 2019 年版，第 15 页。

序，可以挖掘出恐怖主义信息背后隐藏的激进个人、恐怖组织与恐怖主义基地之间的联系模式。因为在暗网信息接触的过程中，暗网黑市的买方和卖方、论坛发帖者与访问者之间可能会在互动过程中识别彼此身份、传授和习得激进意识形态、达成共同的行动意图，而频繁的恐怖主义话题、标签的信息交流凝聚了危险用户关系网络，增加了恐怖主义行为风险。通过一定方法探知恐怖主义线上人际关系结构的潜在模式和发展动态，有助于定位暗网中恐怖组织和追随者个人活动轨迹较为集中的隐蔽社区。

社会网络分析和超链接分析是在已建成的暗网恐怖主义爬虫数据库基础上，以强连接方式确定论坛、黑市使用情况与用户之间相互关系的可视化方法。二者结合能够动态发现和扩展出最大集合的涉恐关联信息，有助于有关机构更加全面、深入地掌握暗网恐怖组织的人员构成及其深度网络活动，为采取措施防范恐怖主义袭击提供有效支持。

具体而言，在暗网站点利用爬虫工具获取恐怖主义相关活动和人员信息后，将涉恐人员之间的互动关系形态用一定的网络结构和关系图例表现出来：暗网黑市的买方和卖方、暗网论坛的发帖者和访问者各以一个节点表示，他们在暗网中的二元活动关系以一条连线表示，[①] 节点间相互指向的图示和度量可以用来映射暗网中的涉恐人员与恐怖组织之间的联系与紧密程度，假定嵌入其中的若干社会关系模式对可能发生的暴恐行动将产生重要影响。

第一，分析出松散结构网络中的核心节点并进行用户画像。高频节点反映暗网中涉恐人员的身份角色和人身危险性，利用节点的计算和排序，分析出位于核心节点的重点人物。根据该重点人物在暗网中的发言内容、交易类型、上网习惯、网络特征、应用场景等多维数据，抽象出一个标签化的用户模型，在多个涉恐黑市或论坛搜索与该模型具有高相似度的用户，循线追踪其真实身份及关系网络，探查是否集结或参与恐怖主义社群。

第二，创建隐蔽社区危险用户交互关系视图。基于多个代表重点人物的关键节点形成强连接的聚类视图，以节点密度、形状、规模和中心点分析恐怖组织或极端主义人员的聚类结构、层次和分布。通过关联分析，找出由危

① Aparna Basu, “Social Network Analysis: A Methodology for Studying Terrorism,” pp. 215-242.

险用户组成、较为稳定的活跃集群，以此勾勒暗网中的恐怖主义隐蔽社区，确定在社区中扮演主要角色并居于组织核心地位的用户。

第三，推演恐怖主义暗网渗透方式与种类。通过聚类社群和隐蔽社区发布信息的时间跨度，分析重点人物的活跃度与某一事件模型发生的同时性、秘密网络结成的时间规律和隐秘社区的扩张速度、变动状态等。提取暗网论坛、黑市集合中所有网页的超链接，包括暗网论坛和暗网黑市中的发帖、评论的外部链接进行聚合列表，发现更多的超链接推送给爬虫数据库，形成网络分析图谱揭示暗网站点相互交互的密切关系。

结 束 语

基于暗网通信与交易的匿名和无痕的特性，打击暗网恐怖主义活动难度较大。而由于网络结构的层级性，以目前的技术手段不可能彻底消除暗网。我们能够采取的应对之策，应当是对表网中用于发现恐怖主义信息的技术原理和分析方法加以改造和升级后，运用于暗网信息的智能分析与深度挖掘，以有效遏制暗网恐怖主义活动。暗网并非“法外之地”，就加强网络安全、筑牢安全防线的现实需要而言，有必要加大资源投入，使更多的研究机构和实务部门能够加入打击暗网恐怖主义的活动，包括利用先进的信息调查和智能数据挖掘技术、运用可视化分析工具等技术手段与方法，侦测恐怖主义线上、线下活动，甄别隐藏在暗网中的恐怖分子及其支持者，遏制暗网恐怖主义滋生和蔓延的趋势。打击暗网恐怖主义活动是一场多兵种配合的综合战，无法凭借单一的技术方法去应对，而应当综合使用多种方式和手段，采取包括但不限于网络战、心理战、军事战等在内的各种措施，加大打击力度，以形成对网络恐怖主义的高压态势。在反恐斗争中提高反恐能力和水平，不断增强应对措施的有效性，不给暗网恐怖主义分子留下可乘之机，在利用好互联网给人类带来的福祉的同时，通过“利剑”“净网”等行动，还互联网一片清朗的空间，增强人民群众的安全感。

[责任编辑：杨 立]