李国灏, 郑昊, 张伟. 基于暗网 3.0 的加密数字货币可追踪方法[J]. 智能计算机与应用,2024,14(9):41-47. DOI:10.20169/j.issn.2095-2163.240906

基于暗网 3.0 的加密数字货币可追踪方法

李国灏, 郑 昊, 张 伟
(南京邮电大学 计算机学院, 南京 210023)

摘 要: 暗网是指需要通过特定配置才能登录的特殊网络。暗网通信突出的特点是匿名性,其导致暗网中滋生大量通过加密数字货币进行的违法犯罪交易,且难以被追踪。本文针对最新的暗网 3.0 版本提出了一种以比特币为代表的加密数字货币可追踪方法,据此设计了集暗网页面爬取、页面信息分类、比特币地址提取、比特币地址追踪、交易信息分析为一体的系统,收集了 13 587 个有效的比特币地址,成功对其中 247 个活跃的比特币地址交易进行了监视,并通过启发式方法获得了交易列表,最后结合明网获取相关关联信息,得到了 35 个较为准确的交易信息表。

关键词: 暗网; TOR; 爬虫; 比特币; 匿名交易
中图分类号: TP391.01 **文献标志码**: A **文章编号**: 2095-2163(2024)09-0041-07

A traceable method for cryptocurrency based on dark Web 3.0

LI Guohao, ZHENG Hao, ZHANG Wei
(School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: The darknet refers to a special network that requires specific configurations to access. Anonymity is the prominent characteristic of dark web communication, which has led to a proliferation of illegal transactions using encrypted digital currency that are difficult to trace. This paper proposes a trackable method for encrypted digital currencies represented by Bitcoin for the latest version of the darknet, Darknet 3.0. Based on our method, a system was designed that integrates dark web page crawling, page information classification, Bitcoin address extraction, Bitcoin address tracking, and transaction information analysis. A total of 13 587 valid Bitcoin addresses were collected, and 247 active Bitcoin addresses were successfully monitored for their transactions. A transaction list was obtained through heuristic methods, and 35 relatively accurate transaction information tables were obtained by combining relevant information from the surface web.

Key words: darknet; TOR; Web crawler; bitcoin; anonymous transaction

0 引 言

互联网为使用者提供了丰富的信息,但并不是其中的所有信息均可被轻易获取。现代网络可以根据信息的深度,分为表网和深网两个部分。日常使用的互联网,如万维网均属于表网的一部分,其仅仅占整个互联网的 10%左右,而剩下的 90%左右属于深网。Fleder M 等[1]指出,深网较平时使用的表网具有更未知的广度和深度,其内容中不含索引与目录,使用者不能直接搜索并访问数据,因此充满了神秘性。本文研究的暗网属于深网的一个子集,其无法通过常用的搜索引擎对其进行查询与访问,需要使用一些特定的配置和非标准通讯协议来进行访问。

建立暗网的初衷是为了保护个人隐私,其独特的匿名技术[2]对建立更加完善的互联网环境有着积极作用。但正是由于其私密性,各种违法犯罪活动在暗网中也得以生存和滋长,如:毒品、儿童色情、人口贩卖和非法数据等。Liang J 等[3]通过研究表明,不仅有个人在暗网中进行违法交易,还有组织性的犯罪集团,而监管部门对于暗网中违法犯罪行为的监管与追踪仍有较大困难。因此,研究暗网中追踪用户的方法可以为监管者、执法者打击暗网中违法犯罪活动提供有效支持。

基金项目: 南京邮电大学 STITP 课题 (202210293101Y)。
作者简介: 李国灏(2001-),男,本科生,主要研究方向: 机器学习,网络信息安全; 郑 昊(2002-),男,本科生,主要研究方向:自然语言处理,数据挖掘。
通讯作者: 张 伟(1973-),男,博士,教授,主要研究方向:智能感知与认知,恶意代码分析,隐私保护。Email: zhangw@njupt.edu.cn
收稿日期: 2023-05-07

研究者在暗网及比特币交易分析方面已经开展了相关研究工作。He X 等[4]分析了比特币交易与地址之间的关联,提出了一种改进的地址变化检测算法,并与原地址变化检测算法进行了比较,通过加入条件约束,使得识别变化的地址更加准确,加快了算法的收敛速度。Zhao Z 等[5]提出了一种增量地址聚类方法,以存储历史状态,并更快地发现比特币的匿名性缺陷。Kim M 等[6]提出了运用可扩展的聚类统计方法进行交易类型与交易规模数据的采集,并构建了比特币用户图以研究其拓扑特性,揭示了比特币网络的匿名性、虚拟性、不可追踪性等多个属性,为追踪加密数字货币提供了基础。Huang Z 等[7]提出了根据比特币地址行为自动分类的工具,将比特币地址的交易转换为地址图结构,并引入图节点压缩技术和图结构增强方法来表征统一的图表示。其精度优于最先进的比特币地址分类器和现有的分类模型。Xiang Y 等[8]讨论了使用图卷积网络、矩阵分解和随机行走等技术对比特币进行去匿名化的优势和挑战,指出在实际操作中存在的困难,但并没有进行实际的操作;Chan W K 等[9]对比特币地址的演变进行了清晰的描述,还解释了一些功能的实现方法。Lee S 等[10]对比特币等加密货币的性质进行了分析,并提出了由聚类、启发式所有权等方法检测和跟踪交易地址,但仅仅获得了比特币交易的非法交易链和比特币流向的总体特征,并未将交易信息与明网信息进行匹配。Tao B 等[11]提出了由地址聚类对比特币网络去匿名化,提出了暗网与表网相结合的分析方法进行信息匹配,但误差较大且数据收集难度巨大,需要结合启发式方法进一步提高结果的准确率。

本文针对暗网中大量存在且难以追踪的非法比特币交易,提出一种集暗网页面爬取、页面信息分类、比特币地址提取、比特币地址追踪、交易信息分析为一体的系统,旨在为相关单位和组织打击暗网中存在的网络犯罪提供有效手段。本文的主要创新点在于实现自动化收集暗网页面信息,对页面进行高效分类;对比特币交易进行追踪,并将暗网与表网结合,获得交易账户的相关信息。

1 相关理论与技术

1.1 匿名通信系统

匿名通信是指采取一定的措施来隐蔽通信流中的通信关系,以实现窃听者难以获取或推知通信双方的关系及内容,其目的是为了保护个人通信隐私。匿名通信系统主要由提供加密服务的节点组成,利用内容加密和网络混淆等多种技术来提供匿名通信服务。暗网是匿名通信系统的表现形式之一,其利用隐藏服务机制,使得暗网中的非法活动无法被有效追踪。

洋葱路由网络[12](The Onion Router,Tor)作为最典型的匿名通信系统之一,其发展在近年呈现出迅猛态势。Tor 网络由客户端、洋葱代理、洋葱路由节点、目录服务器及网桥服务器组成。在通信过程中,消息被一层一层加密成犹如洋葱一样的数据包,保证目的地可以获得原始消息,并且整个通信过程难以被追踪。

除 Tor 外,还有许多匿名通信系统,比较典型的有大蒜网络[13](Invisible Internet Project,I2P)、自由网[14](Freenet)及零网[15](ZeroNet)等。I2P 基于 P2P 通过不同的隧道,将中间节点与目标节点分隔出来。但与 Tor 的洋葱路由不同,I2P 的核心是大蒜路由,将传输的原始数据拆散为加密数据包通过多条隧道交叉疏散传递,经过传输隧道层层解密后到达目标节点。Freenet 使用基于密钥的路由协议,通过去中心化的分布式数据存储来保存和传递信息,仅通过传递内容请求并在不知道完整文件内容的情况下,将请求发回的中间计算机进行连接,重点在于言论自由和匿名性。ZeroNet 则采用 Python 构建,通过公钥来识别的点对点用户网络,其私钥允许站点所有者签署和发布更改并通过网络传播。ZeroNet 默认不是匿名的,但其支持通过 Tor 网络路由流量。

1.2 暗网 3.0

暗网服务器的网址以 onion 为顶级域名,并由其公钥派生出的字符组成。V2 版本的暗网洋葱地址为 16 个字符长,但在 2021 年底,暗网格局发生了巨变,暗网由 V2 版本更新为 V3 版本。V3 版本的暗网洋葱地址为 56 个字符长,从而导致许多旧网站无法进行访问。更新的 v3 版本较 v2 版本,签名算法从 SHA1/DH/RSA1024 升级到了 SHA3/ed25519/curve25519;升级的 Tor directory protocol 具有更高的安全性;洋葱地址更换成 sha3,提高了生成相同地址的难度;改进的目录协议保证了目录服务器更小的被攻击面和更少的信息泄露。

1.3 比特币

比特币作为加密数字货币的典型代表,是暗网市场交易的主要媒介。中本聪于 2008 年首次提出比特币的概念,十几年内比特币飞速发展,价格一路

飙升，引起了多国政府的关注。不同于普通货币，比特币不由货币机构发行，而是由大量计算产生。Vranken H[16] 说明了比特币的相关原理，其运用 P2P 网络众多节点构成的分布式数据库确认与记录所有交易，并运用密码学原理来保证流通过程中的安全性。正因为加密方式特殊，导致交易中存在大量非法的交易行为。对比特币进行追踪分析，可以为执法者打击暗网中的犯罪行为提供思路。

1.4　文本处理技术

词频－逆文档频率算法（Term Frequency－Inverse Document Frequency，TF－IDF）[16] 是一种用于信息检索与文本挖掘的常用加权技术，用以评估一个字词对于一个文件集的重要程度，其由词频（Term Frequency，TF）和逆向文件频率（Inverse Document Frequency，IDF）组成。字词的重要性随着其在文件中出现的次数成正比增加，但同时会随着其在语料库中出现的频率成反比下降。

文本排序算法（TextRank）[18] 是一种基于图的用于关键词抽取的排序算法。其利用一篇文档内部的词语间的共现信息便可以抽取关键词，能够从一个给定的文本中抽取出该文本的关键短语、关键词组，并使用抽取式的自动文摘方法抽取出该文本的关键句。其基本思想是将文档看作一个词的网络，该网络中的链接表示词与词之间的语义关系。

相比于 TextRank，快速自动关键字提取算法（RAKE）是一种对英文文本处理效果更好的算法。RAKE 提取的是关键短语，且倾向于较长的短语关键点，将短语中每个单词的共现关系考虑了进去。

2　系统分析与设计

2.1　系统整体框架

结合第一章所论述的理论和方法，针对暗网中存在大量非法交易问题，本文设计了一套自动收集并分析暗网交易和比特币流向的系统，系统流程如图 1 所示。系统主要实现了 3 个目标：爬取最新版本的 Tor 暗网域名，获得并分析有效页面信息；挖掘比特币地址并关联；交易账户信息挖掘。

系统采用了现有的一些暗网域名爬取及页面信息分析技术，主要工作集中在比特币地址的挖掘与关联以及交易账户信息的挖掘上。具体内容如下：

（1）暗网网页信息的自动处理。自然语言处理中的文本处理针对的大多是规范的文本，而暗网页面中存在大量的标签、转义符、特殊符号等干扰因素。

（2）涉及非法交易的比特币地址收集。大部分暗网页面都是多级页面，在进行比特币交易之前需要验证、登录等一系列复杂操作。

（3）非法交易账户信息的挖掘。暗网中信息难以收集，与明网的信息差别较大，且暗网中开展违法犯罪的嫌疑人具备较高的防范意识，导致具体信息难以关联到交易账户。

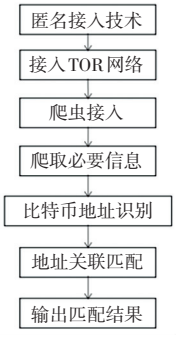

图 1　系统流程

Fig. 1　System process

2.2　基于 onionscan 的数据收集

2.2.1　暗网爬虫接入

由于地域限制，Tor 网络的接入必须使用 vpn 并使用其内置的 obfs4 网桥。使用的 obfs4 网桥为：obfs4 213.136.46.231:1213 C477DFFA6C110702513900CD9772FEDBC9417D68 cert=kDJztEJ/t/zonemEUe0YF/0yBmkZUoD32vRIyyrrLClvAaTDEx10XTlVtrcODKk8gACYOA iat-mode=0，而 python 爬虫请求访问网站时无法使用该功能。为了消除地域影响，系统租用了一台服务器。Tor 在数据传输时使用的是 socks5 协议，而 python 中的 requests 爬虫模块使用的是 http 协议，所以需要在爬虫程序中设置 socks5 代理以连接 Tor 网络。使用 curl 命令 curl --socks5-hostname 127.0.0.1:9150 xxxxx.onion 以快速测试是否已经通过 socks5 代理与 Tor 网络建立连接。

搭好配置，使用 MobaXTerm 工具登录并建立会话，选择 SSH 协议进行链接。成功登录后，设置防火墙与 Tor 客户端，修改 Tor 配置文件，接入经 Arisoy M V 等[19] 比较选出的性能优秀的 python 爬虫。暗网爬虫接入成功，则可以获取页面源代码。

2.2.2　暗网域名收集

系统通过明网的门户网站，爬取所需的暗网网址。利用谷歌的搜索，选取 5 个明网门户网站（https://darkweblinks.com/、https://thehiddenwiki.org/、https://darkweb-sites.org/、https://torsites.org/和 https://www.deeponionlinks.com/）爬取。由于网站的网页结构相似度较高，暗网域名存在于特定标签中，故运用 BeautifulSoup 解析页面，获取所需

暗网域名,存入列表。在收集到的域名基础上,使用 socks5 代理,由代理服务器解析域名,在成功请求后,利用 text 函数获取源代码。接着使用正则表达式匹配源代码中的域名,利用字符串长度的特征筛选有效暗网域名,存入新列表。重复上述过程,系统最终得到 8 564 个暗网域名。

2.2.3 爬取页面处理

暗网域名得到后,系统使用相同的爬虫获取其所对应的页面源代码。由于部分暗网页面的存在周期极短,在二次爬取时,页面已经不可访问,因此这类页面源代码无法获取。同时,在爬取成功的页面中,有部分页面是无效页面或为验证人机页面等。这些页面不能提供进一步分析的数据,则进行清除。剩余页面中少部分存在编码格式混乱,将这类页面移入其余位置单独保存。经过系统页面处理,得到了 7 896 个地址,涵盖枪支、毒品、儿童色情、投资、赌博、邮箱、黑客等多方面的交易。

表 1 展现了爬取到的暗网域名中的 3 个典型域名,其分别涉及不同的交易内容。其中,第一个地址涉及色情交易,其中含有 lolita 这类识别性强的关键词;第二个地址涉及比特币交易,其中含有 btc、transaction 等关键词;第三个地址内容主要是枪支交易,关键词中均为枪支名称。在后续的工作中,选取涵盖各类交易的比特币地址进行追踪分析,以保证实验样本的涵盖面。

表 1 爬取的部分暗网域名

Table 1 Partially crawled dark web domain names

Address	Key Words
qegmne5isc56rs7vhqk7tm6bxmuehskkh60br63pi6dkke4jqkgbicid. onion	views, pussy, underage, video, lolita
gtyunnpqvdeuxtoxzge4wmx5gfbaa5ys3j337ctfdab7hcqrw3d3c2yd. onion	transfer, btc, transaction, link
gunsmohvegmcwbt5si75li22uq5cxoseuhtd2huqy3sxym2jq7k7usid. onion	cart, ak-, rifles, add, guns

2.3 系统功能模块设计

由于暗网的匿名性,交易双方的比特币地址通常是匿名的,不会直接显露出来。因此,本文通过分析交易数据、跟踪交易行为和核对真实交易获取更加丰富的信息来支撑后续的分析过程。

(1)分析交易数据:通过分析暗网上的交易数据,可以确定哪些地址之间发生了交易,尽管地址本身是匿名的,但通过交易之间的联系可以推断出交易双方的可能身份。有多家著名的比特币交易所提供交易信息查询服务,如火币全球专业站。在地址丰富列表中,可以看到比特币地址数量分布图、比特币地址余额分布图。通过地址数量分布图,可以看到地址与货币持有比例对应的比特币数量;通过地址余额分布图,可以看到比特币地址余额的总量和比例,以及相应的数字。通过查询某个比特币地址,可以查询到该地址的余额、交易数量、交易记录、交易编号、交易相关地址等相关信息。表 2 展示了其中一个比特币地址账户的相关信息,可以获得该比特币账户的区块高度、出块时间、输入输出具体金额等关键信息。系统从深度爬取多级页面收集到的 13 587 个有效比特币交易地址中选取了 561 个活跃度较高、交易类型不同的地址进行监视。

表 2 比特币账户信息

Table 2 Bitcoin account information

交易信息	具体内容
区块高度	780797
出块时间	2023/3/14 23:48:39
输入	4 632.126 702 67 BTC
输出	4 632.126 330 07 BTC
手续费	0.000 372 60 BTC
金额	4 632.126 702 67 BTC

(2)跟踪交易行为:通过监视选定的地址并记录交易行为,可以收集关于该地址的更多信息,如哪些地址曾经向该地址发送过比特币或从该地址接收过比特币。表 3 展示了一个比特币账户的相关交易信息,该账户接受由同一比特币地址输入的 3 笔交易,后将所有的比特币转出给两个不同的账户。输入与输出金额之差为扣除的手续费。

表 3 比特币账户交易信息

Table 3 Bitcoin account transaction information

输入	BTC	输出	BTC
1KsoefB4C85xUrnvGmQ147obSRod6tnQjQ	4 614.771 806	1S5SyH8RLCboWoHRdbhc8Vh17hfTyUDyN	5.051 300 87
1KsoefB4C85xUrnvGmQ147obSRod6tnQjQ	17.354 576	1yBLkKHpHqkpjphys2yS5KswAkFKe3ffp	4 627.075 029
1KsoefB4C85xUrnvGmQ147obSRod6tnQjQ	0.000 320 8		

(3)核对交易记录：使用区块链浏览器可以查看所有比特币交易的公共记录，常用的区块链浏览器有：莱特币区块链浏览器、以太坊区块链浏览器、达世币区块链浏览器和 BTC 区块链浏览器。要关注特定的交易，可以通过查看区块链浏览器中的相关数据，找到该交易涉及的地址，从而确定交易双方的比特币地址。此外，为了更好地匹配交易双方的信息，不仅要关注交易的账本记录，还要关注交易金额、交易开始时间和结束时间。当金额与时间具有较强的一致性时，便可以认为已经确定了交易发生的双方，从而完成较高准确度的比特币地址匹配。表 4 通过比对两个地址的交易时间、交易金额、账户地址等信息，来确保交易在两者之间已经发生。系统最终从 561 个活跃度较高的地址中成功监视了 247 个比特币地址，并获得了准确的交易信息。

表 4　对比交易信息

Table 4　Comparison of transaction information

比较项目	输入	输出
地址	1KsoefB4C85xUrnVG	af66a7ac0cae28b2b43183
	mQ147obSRod6tnQjQ	569adfa9284b562ddb7773
		97cb8237c79a8db0808d
区块高度	780 797	780 797
交易时间	2023/3/14 23:48:39	2023/3/14 23:48:39
交易金额	4 614.771 806	4 614.771 806
手续费	0.000 372 60 BTC	0.000 372 60 BTC

在数据收集模块已经挖掘出了大量暗网网址与域名，分析了其中所蕴含的信息，并由上述 3 种方法分析相关交易信息，寻找所需要的交易双方的比特币地址。

需要注意的是，虽然可以通过以上 3 种方法在某种程度上识别交易双方的比特币地址，但交易双方的真实身份仍然无法确定，因为这些地址仍然是匿名的，仅能得到交易相关账户的列表。要实现高程度的信息匹配，需要通过启发式方法并结合表网分析将比特币地址与其现实身份相匹配。

2.4　与表层网结合分析

比特币地址本身并没有直接关联现实身份的功能，比特币的设计是基于去中心化和匿名性的原则，每个比特币地址只是一个由随机数字和字母组成的字符串，并不包含任何个人身份信息。

已知市面上所交易的比特币基本是从比特币交易所中交易而来，个人挖掘的比特币也会在交易所被交易。从这个角度入手，将比特币地址与现实身份相关联。例如：一个用户在某个交易所注册账户并购买了比特币，这个交易所会收集这个用户的个人身份信息，并将其交易记录与其身份相关联。如果这个交易所透露了用户信息，或者通过攻击或爬取的方式获得交易所的交易信息，那么该用户的比特币地址可能会被关联到他的现实身份。

此外，如果在公开场合或社交媒体上透露了自己的比特币地址，或者与其比特币地址相关的信息，这些信息也可用于关联用户的现实身份。

在 Lin Y J 等[20]研究基础上，构建了一个相关性度量系统，将得到交易双方的比特币地址与现实明网的信息关联，或者与交易所的信息匹配，便可以从交易双方的比特币地址找到其对应的现实身份。按照 Kanemura K 等[21]的方法，从收集到的账户信息中选取一些特征作为度量，熵值作为反映离散程度的重要指标，计算明暗网账户的信息熵来反映关联性。具体过程如下：

1)数据标准化

由于明暗网账户之间的差异较大，采用最大值最小值法对数据进行标准化，公式如下：

$$x_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \tag{1}$$

式中 X_{ij} 表示暗网账户第 j 个信息的度量。

2)计算第 j 个度量的信息熵：

$$e_j = -k\sum_{i=1}^{n} p_{ij}\ln(p_{ij}) \tag{2}$$

式中：$k = \frac{1}{\ln(n)}$，$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$。

3)计算每个度量的权重：

$$w_j = \frac{1 - e_j}{\sum_{j=1}^{m}(1 - e_j)} \tag{3}$$

式中 m 是对度量的计数。

4)生成结果

对每个度量进行加权平均：

$$u_i = \sum_{j=1}^{m} x_{ij} \cdot w_{ij} \tag{4}$$

若生成的结果值较大(大于等于 100)，则可认为暗网明网信息关联性极强，是对应的账户信息。在前面得到准确交易信息的比特币 247 个地址中，经过度量计算后，有 35 个地址结果值符合要求，即成功进行了信息匹配。

3　实验结果分析

实验通过租用境外服务器解决了 Tor 连接的问

题,并通过修改爬虫的 http 协议成功实现将爬虫接入暗网,实现了暗网页面的自动化收集。由从明网上获取的 16 个暗网门户网站,通过扩大爬取,收集到 8 564 个有效的暗网地址。再对爬取的暗网地址进行剔除、筛选,最终得到 7 896 个用于后续分析的地址。

由图 2 可见,大部分比特币地址都涉及交易类,涵盖枪支、毒品、儿童色情、投资、赌博、邮箱、黑客等多个方面的交易。而图中单独列出的黑客类、毒品类和枪支类,是指介绍相关商品,而不涉及交易的地址。此外,门户类、色情类和谋杀类也占有一定的比例,在暗网中屡见不鲜。

在这些页面中,通过深度爬取多级页面,并进行验证,收集到了 13 587 个有效的比特币交易地址。通过在区块链浏览器上搜索相关地址,系统选取了 561 个活跃度较高、交易类型不同的地址,监视其中出现频率较高的交易,来观察交易的账本和交易时间,得到交易双方地址。系统成功对其中 247 个比特币地址的交易进行了监视,得到了每个交易对应的地址列表和交易流向表。

最后,在暗网与表层网上进行身份匹配,得到了 35 个准确的交易双方信息,如名字、身份证号、IP 地址等关键信息,为执法者对犯罪行为进行追踪提供了有力的信息切入点。表 5 展示了部分信息匹配的结果,大部分交易可追踪到的信息只有一类,暴露信息大于一类的交易仅占总交易数量的 12%,占比较小。在后续的工作中,要得到同一账户不同类型的交易信息,需要突破明暗网之间的信息差,以扩大信息量,进行更深层次的信息搜寻与关联。

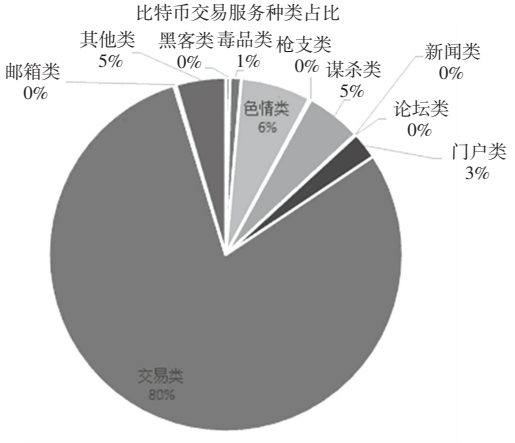

图 2 比特币地址参与服务类型占比

Fig. 2 Proportion of service types associated with Bitcoin addresses

表 5 部分信息匹配结果

Table 5 Partial match results of information

交易哈希	呢称	ID card	IP 地址	facebook
a6cfcca8af04e83d8005c88ecb98ffae8 46989fa21c2b277fd0fbd195fb5b260	lachin filand			
bc1quq29mutxkgxmjfdr7 ayj3zd9ad01d5mrhh8912		2217455901		
15uyKEJ9HRC7K6Go2 XMK97Le4xgkuCGrjaA	alison%M			
13VpJb5XTLZ97D2Ff P5q7jmqYwrYaPnd1v	baxsende may			
be1qtxujlhquy68e6h8rjw g2lns5pechjnx3fb7x77			66. 11731. 255	sdwc569zwcfrg
3031a6296580435c4a6elald30298e2f 3a9c3d512b51133b2289e4a9a6286acb	Odnjcx			

4 结束语

本文对暗网中非法交易进行可行性追踪研究,实现了暗网中交易信息的爬取、筛选和收集,获得相关比特币账户交易流向图,并创新性地提出将暗网中的交易信息与明网中的账户信息进行匹配这一关键步骤,进一步实现对加密数字货币的追踪,为执法者打击暗网中的违法交易提供了技术支持。

参考文献

[1] FLEDER M, KESTER M S, PILLAI S. Bitcoin transaction graph analysis[J]. arXiv preprint arXiv:1502. 01657, 2015.

[2] HONG S，KIM H. Analysis of Bitcoin exchange using relationship of transactions and addresses [C]//Proceedings of 2019 International Conference on Advanced Communication Technology (ICACT). IEEE，2019：67-70.

[3] LIANG J，LI L，ZENG D D. Bitcoin exchange addresses identification and its application in online drug trading regulation [C]// Proceedings of the 23rd Pacific Asia Conference on Information Systems. 2019.

[4] HE X，HE K，LIN S，et al. Bitcoin address clustering method based on multiple heuristic conditions[J]. IET Blockchain，2022，2(2)：44-56.

[5] ZHAO Z，WANG J，SHI K，et al. Improving address clustering in bitcoin by proposing heuristics [J]. IEEE Transactions on Network and Service Management，2022，19(4)：3737-3749.

[6] KIM M，LEE J，KWON H，et al. Get off of Chain：Unveiling dark web using multilayer bitcoin address clustering [J]. IEEE Access，2022，10：70078-70091.

[7] HUANG Z，HUANG Y，QIAN P，et al. Demystifying bitcoin address behavior via graph neural networks [J]. arXiv preprint arXiv：2211.14582，2022.

[8] XIANG Y，LEI Y，BAO D，et al. BABD：A Bitcoin address behavior dataset for pattern analysis [J]. IEEE Transactions on Information Forensics and Security，2024，19：2171-2185.

[9] CHAN W K，CHIN J J，GOH V T. Evolution of Bitcoin addresses from security perspectives [C]// Proceedings of 2020 15th International Conference for Internet Technology and Secured Transactions. IEEE，2020：1-6.

[10] LEE S，YOON C，KANG H，et al. Cybercriminal minds：An investigative study of cryptocurrency abuses in the dark web[C]// Proceedings of 26th Annual Network and Distributed System Security Symposium. IEEE，2019：1-15.

[11] TAO B，DAI H N，WU J，et al. Complex network analysis of the bitcoin transaction network[J]. IEEE Transactions on Circuits and Systems II：Express Briefs，2021，69(3)：1009-1013.

[12] SCHRINER J. Weaving the dark web：Legitimacy on freenet，tor，and i2P[J]. Internet Histories，2019，3(3-4)：388-390.

[13] DINGLEDINE R，MATHEWSON N，SYVERSON P F. Tor：The second - generation onion router [C]// Proceedings of USENIX Security Symposium. IEEE，2004：303-320.

[14] ZANTOUT B，HARATY R. I2P data communication system [C]//Proceedings of ICN. 2011：401-409.

[15] WANG S，GAO Y，SHI J，et al. Look deep into the new deep network：a measurement study on the ZeroNet[C]// Proceedings of Computational Science - ICCS 2020：20th International Conference. IEEE，2020：595-608.

[16] VRANKEN H. Sustainability of bitcoin and blockchains [J]. Current Opinion in Environmental Sustainability，2017，28：1-9.

[17] BAFNA P，PRAMOD D，VAIDYA A. Document clustering：TF-IDF approach [C]// Proceedings of 2016 International Conference on Electrical，Electronics，and Optimization Techniques (ICEEOT). IEEE，2016：61-66.

[18] MIHALCEA R，TARAU P. Textrank：Bringing order into text [C]//Proceedings of the 2004 Conference on Empirical Methods in Natural Language. 2004：404-411.

[19] ARISOY M V，KÜÇÜKSILLE E U. Performance comparison of TOR hidden service crawlers [J]. Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi，2019，6(2)：147-161.

[22] LIN Y J，WU P W，HSU C H，et al. An evaluation of bitcoin address classification based on transaction history summarization [C]// Proceedings of 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE，2019：302-310.

[21] KANEMURA K，TOYODA K，OHTSUKIT. Identification of darknet markets' bitcoin addresses by voting per - address classification results[C]// Proceedings of 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE，2019：154-158.