RUSI
www.rusi.org
Royal United Services Institute
for Defence and Security Studies

HUMAN COGNITION

Global Research Network on Terrorism and Technology: Paper No. 6

Mapping the Jihadist Information Ecosystem

Towards the Next Generation of Disruption Capability

Ali Fisher, Nico Prucha and Emily Winterbotham

Online disruption efforts generally aim to reduce the availability of jihadist content. Yet, the speed and agility of jihadist movements online – a multiplatform approach which a co-author of this paper has previously described as a 'swarmcast' – has allowed groups to evolve in response to disruption efforts and find new ways to distribute content.[1]

This paper describes a model of the flow of users between social media platforms and surface web pages to access jihadist content, using data obtained through innovative collection methods.[2] The model provides an approximate picture of the jihadist information ecosystem and how multiple platforms are used to disseminate content.

Key Findings

- The jihadist information ecosystem is a large and complex network, connecting a vast array of platforms across the surface and dark web.
- Despite claims to the contrary, jihadist content is widely accessible via mainstream social media and the surface web. As of 2019, while jihadists prefer to communicate with core supporters on Telegram, they also use Telegram to coordinate efforts to exploit ('raid' in their terms) other platforms to achieve greater reach and build resilience for the jihadi information ecosystem.
- Adopting a multiplatform communication paradigm (MCP) rather than focusing of individual platforms will be key to developing next-generation approaches to online disruption and content removal.
- This research built a model depicting the flow of internet traffic to content shared by jihadi groups. The model shows different online platforms fulfil different purposes for jihadist groups, and this makes their communications harder to disrupt. As a result, when developing disruption approaches it is essential to differentiate between how links to material are shared and where the specific content is actually stored online.
- While previously online jihadi activity was focused on posting content directly on major platforms, the research shows major platforms are now often used to share URLs instead to facilitate access to content stored elsewhere.
- The major platforms are therefore being used as 'beacons' directing users to the material. Over half of known sources of traffic in the data came from just three platforms – Facebook, Telegram and Twitter. At least 50% of the actual content captured in the model is then stored on specific websites run by jihadist or theologically aligned groups. As these are smaller and obscure servers this takes time to locate and remove.

1. Ali Fisher, 'Swarmcast: How Jihadist Networks Maintain a Persistent Online Presence', *Perspectives on Terrorism* (Vol. 9, No. 3, 2015), pp. 3–20.
2. This research focuses on jihadist groups, but the methodology could be applied to other violent groups.

- Major platforms search for jihadist content on their servers using techniques such as image hashing – but in many cases, these platforms are being used as 'beacons' to share URLs but the content itself is stored on other platforms. This makes the content much harder to detect by the original domain owner.
- Jihadist movements disseminate some of their content through texts, which tend to be uploaded to a different and diverse group of platforms from those used to store audio-visual content. Texts often remain undetected due to the majority being in Arabic and because PDFs or Microsoft Word documents can be uploaded via more platforms than audio-visual content.
- Jihadist networks also share a wide range of extremist Salafist documents. This type of content is often not removed due to the complexity of defining what is actually jihadi content as opposed to some extreme Salafist material. This type of content is at times endorsed, reshared and in some cases re-published with Daesh (also known as the Islamic State of Iraq and Syria, ISIS) media organisation logos but remains available.

Summary of Recommendations

As this research has highlighted jihadist organisations' increasingly complex MCP, the recommendations focus on the steps needed to develop a new generation of approaches rather than specific ways to fine-tune the current 'whack-a-mole' disruption paradigm.[3]

The technology sector should embrace a multiplatform approach in relation to URL and shortlink reporting at greater scale and develop a robust shared awareness of URLs/shortlinks leading to jihadist content across the full range of platforms. While some companies have already successfully deployed this approach, based on services provided by organisations, such as Human Cognition, the research shows many others could benefit if individual companies or the sector more broadly developed a mechanism to use these services effectively.[4]

3. Suspension constitutes a repressive rather than preventive counter-radicalisation measure. It can remove accounts but cannot stop 'returners', who frequently re-establish new accounts post-removal, known as the 'whack-a-mole' effect. See, Jessica Stern and J M Berger, *ISIS: The State of Terror* (London: William Collins, 2015); Rachel Briggs and Sebastien Feve, 'Policy Briefing: Countering the Appeal of Extremism Online', Institute for Strategic Dialogue, 2014. The evolution of jihadi strategy to avoid 'whack-a-mole' approaches is discussed in Fisher, 'Swarmcast'.
4. Human Cognition has previously provided parts of the data in this report to some governments and platform providers, either through the near real-time approach discussed here or for specific research projects.

The tech sector and researchers should also focus on the functions of each platform within the ecosystem. This would facilitate targeted disruption strategies, as recognising how a platform is used within the overall effort to distribute extremist material and hate speech is key to enhancing disruption. Researchers should also factor platform function in the design of research projects.

The ability to penetrate digital networks where jihadist groups initially distribute new URLs will increase the speed of detection as it effectively creates an early warning system. Platform owners should develop or acquire an early warning system to detect new URLs and shortlinks that store or point to jihadist content, through infiltration of online jihadist networks or tracking of users sharing jihadist content, within the constraints of existing and proposed privacy legislation, and while ensuring duty of care to employees.

Technology companies and internet referral units should recruit analysts with Arabic language skills and knowledge of jihadist and extreme Salafist texts. Human verification processes alongside machine learning techniques should be considered best practice.

As governments propose new legislation on privacy protection (such as that in the EU)[5] policymakers must ensure that their attempts to protect privacy and to encourage platforms to tackle extremist content are not contradictory. As governments threaten to fine tech companies for failing to remove content quickly from their platforms, they cannot continue to work with researchers who have been actively posting jihadist content on the same platforms.

Introduction

Jihadist movements adapt rapidly to changes in their operating environment online and exploit an approach previously labelled by one co-author of this paper as a 'swarmcast', characterised by speed of distribution, agility, and resilience of the network structure.[6] This online struggle for survival dictates an evolutionary logic and constant innovation, leading to the development of a complex information ecosystem that spans multiple platforms.[7] By storing

5. The European Parliament and the Council of the European Union, 'Regulation (EU) 2016/679 of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)', *Official Journal of the European Union* (L119/1, 4 May 2016).
6. Fisher, 'Swarmcast'.
7. Ali Fisher, 'Netwar in Cyberia: Decoding the Media Mujahidin', CPD Perspectives, Paper 5, October 2018, <https://www.uscpublicdiplomacy.org/sites/uscpublicdiplomacy.org/files/Netwar%20in%20Cyberia%20Web%20Ready_with%20disclosure%20page%2011.08.18.pdf>, accessed 10 July 2019.

content on some platforms and directing users to it from other platforms, terrorists make their content (including extremist Salafist content they wish to be associated with) harder to detect and remove.[8]

The dominant tactical approaches used to disrupt this activity have been **suspension** (disrupting dissemination by removing individual websites and social media accounts with the intention of reducing the number of users that jihadist groups could reach) and **deletion** (searching for, identifying and removing individual pieces of content – increasingly leaning towards the use of automated classifiers and machine learning).

While each approach has had considerable impact, they involve an extended game of what has been termed 'whack-a-mole': these approaches cannot prevent people from re-establishing new accounts post-removal.[9] This suggests that a third approach is needed that moves beyond the tactical disruption of individual users or pieces of content, surveys the whole jihadist information ecosystem (and not just a few well-known platforms), and addresses the significant proportion of jihadist content in Arabic.

Research Objectives

This research was designed to produce a network map of the traffic between the online platforms and websites used by jihadists to construct their information ecosystem, and thereby provide a more nuanced understanding of the ways in which jihadist content is shared to assist future online disruption efforts.

The project aimed to answer the following research questions:

1. How do jihadi groups and terrorist organisations share content across a wide range of online platforms?
2. What roles do the different platforms play within the information ecosystem?
3. What are the implications for current disruption and removal efforts?

8. Nico Prucha, 'IS and the Jihadist Information Highway – Projecting Influence and Religious Identity via Telegram', *Perspectives on Terrorism* (Vol. 10, No. 6, 2016), pp. 48–58.
9. See, for example, Stern and Berger, *ISIS*, pp. 136–38, 142–43.

Methods and Data Sources

The research drew on a comprehensive existing dataset of jihadist, including Daesh, communication in Arabic based on messages posted in channels and chat groups in Telegram and collected via BlackLight from 1 January 2016 to 31 December 2018. BlackLight is a cloud-based monitoring system built by Human Cognition.[10] To see how jihadist groups exploit multiple online platforms and websites in what is arguably Daesh's native language (Arabic), the authors used BlackLight to access three years of human-verified jihadist Telegram channels archived in near real-time. A fluent Arabic linguist and subject-matter expert verified each channel as jihadist. All data, diagrams and graphs below are based on this data.

This means that the data captures theologically aligned sources from which jihadist groups draw inspiration and known terrorist content. The rationale for that is that jihadist groups are taken to include those that are defined internationally as terrorist groups and those that are illegal in some parts of the world but not in others. Given the current absence of a globally agreed definition of terrorist and accepted international norms regarding what is extremist and/or terrorist content, it is important to analyse multiple sources of data to understand how the jihadist movement, and groups such as Daesh specifically, navigate the online world.

Using shortlinks extracted from the BlackLight data archive, the researchers developed a model to represent the flow of users between social media platforms and pages on the surface web to access known jihadist content. The result is the largest multiplatform model of the jihadist information ecosystem yet described. In the absence of any legal or practical method of measuring all traffic across the internet, this model provides an approximation of the jihadist information ecosystem.

Preliminary findings were presented at an expert workshop, attended by academics and representatives of the Global Internet Forum to Counter Terrorism (GIFCT). Their feedback has been incorporated into this final paper. As a result of feedback from GIFCT members, the authors examined a subset of URLs to assess whether they were still accessible. The results of this exercise can be found below under Longevity of Content.

10. Data collection was conducted using BlackLight – a product of Human Cognition, which has continuously monitored human-verified jihadist Telegram channels in near real-time for over two and a half years. BlackLight has archived the posts, including all URLs shared over that period. This provided the archive from which shortlinks were identified.

What are Shortlinks – and How Does This Help?

Shortlinks allow any user to share a convenient short URL instead of sharing a long and unwieldy weblink. Some of the most familiar shortlinks are Twitter's t.co, Google's goo.gl, WordPress's wp.me, and bit.ly. Shortlinks often have access to the platform on which users originally clicked the link. Some shortlink services report this information, along with how many times a link has been clicked.

Using social network analysis (SNA), the data aggregated from shortlinks was built into a model of the information ecosystem that represents the traffic between the platforms where users click links and the location to which they are redirected. For example, a user may come across the shortlink on Facebook, click it and be redirected to the content on YouTube – this would count in traffic analysis as Facebook being the source of traffic and YouTube being the target.[11]

Why Focus on Arabic?

While many reports focus on social media accounts and sources that use English, [12] Arabic is the primary language of most Salafi-jihadist groups such as Daesh and Al-Qa'ida. Daesh disseminates Arabic texts knowing that such content reaches their Arabic-speaking target audience and bypasses the vast majority of the non-Arabic speaking counterterrorism policy officials,

11. This data is provided at domain/subdomain level only (for example, m.Facebook.com) rather than individual page level. Services do this to protect the anonymity of individual users, among other reasons. This means the project does not handle the data of the users clicking the shortlinks nor the specific social media users sharing them.
12. Of the many articles which focus on English magazines, see, for example: Julia Musial, '"My Muslim Sister, Indeed You Are a Mujahidah" – Narratives in the Propaganda of the Islamic State to Address and Radicalize Western Women. An Exemplary Analysis of the Online Magazine Dabiq', *Journal for Deradicalization* (No. 9, 2017), pp. 39–100; Miron Lakomy, '"One of the Two Good Outcomes": Turning Defeats into Victories in the Islamic State's Flagship Magazine *Rumiyah*', *Terrorism and Political Violence* (15 August 2018), doi:10.1080/09546553.2018.1506335; Amy-Louise Watkin and Seán Looney, '"The Lions of Tomorrow": A News Value Analysis of Child Images in Jihadi Magazines', *Studies in Conflict & Terrorism* (Vol. 42, No. 1–2, 2019), pp. 120–40; Peter Wignell et al., 'A Mixed Methods Empirical Examination of Changes in Emphasis and Style in the Extremist Magazines *Dabiq* and *Rumiyah*', *Perspectives on Terrorism* (Vol. 11, No. 2, 2017), pp. 2–20; A G Reed and Haroro J Ingram, 'Exploring the Role of Instructional Material in AQAP's Inspire and ISIS' Rumiyah', Europol, 2017; Stuart Macdonald et al., 'A Study of Outlinks Contained in Tweets Mentioning "Rumiyah"', Global Research Network on Terrorism and Technology: Paper No. 2, RUSI, June 2019.

academic analysts and commentators. Since the 1980s, jihadist groups have issued over 300,000 pages in Arabic promoting their brand of theology to justify violent jihad.[13] In addition, contemporary Salafi-jihadist material references elements of the rich 1,400-year-old tradition of Islamic writings with terrorist groups sharing the most supportive Salafist writings alongside their own, more recent, products. Exploring content in Arabic is therefore key to understanding how jihadist groups exploit multiple online platforms, and what they share on them.[14]

The data spans two and a half years of continuous monitoring of core jihadist social media via BlackLight. From the BlackLight archive, 1,648 unique shortlinks lead to confirmed jihadist content. On average, each link was clicked over 1,000 times, producing a total of over 1.8 million clicks with a mean of 1,097.37 and median of 249 clicks. These 1.8 million clicks originate from over 150 countries, territories and regions.

Findings

Targets of Traffic

Shortlink services connect the new shortlink to the original long link that was shortened. By aggregating the domain-level data from each of the long links, the platforms where jihadist content is frequently stored can be identified.

13. Seth Cantey and Nico Prucha, 'Reading the Lips of Jihadism: Islamic State Theology, Offline Operations and Online Agitprop', manuscript, forthcoming article for the edited series Studying Jihadism by Rüdiger Lohlker, Vienna University Press, forthcoming 2020.
14. Reuven Paz, 'Reading their Lips: The Credibility of Jihadi Web Sites as "Soft Power" in the War of the Minds', December 2007, <https://www.semanticscholar.org/paper/Reading-Their-Lips-%3A-The-Credibility-of-Jihadi-Web-Paz/60bbc3d355e2d5424d10aefc915368fd03761359>, accessed 15 July 2019.

Figure 1: Targets of Traffic

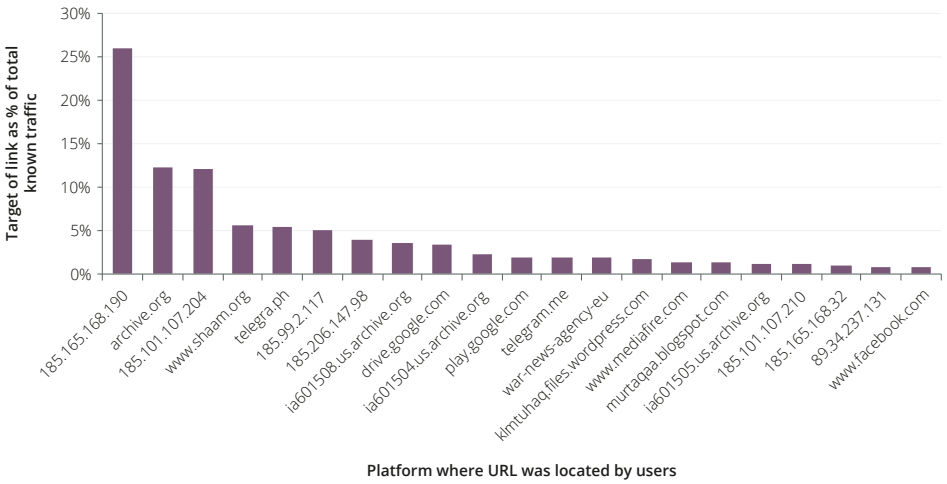


Figure 1 shows the locations to which traffic was directed. Over 50% of known traffic went to self-hosted websites, such as 185.165.168.190. These sites can be accessed via domain names or IP addresses.[15] A further 18% connect to archive.org servers, while approximately 14% connect to other mainstream platforms, such as Facebook or Google. Shaam Network (shaam.org) is a non-jihadist site mainly concerned with news related to Syria.

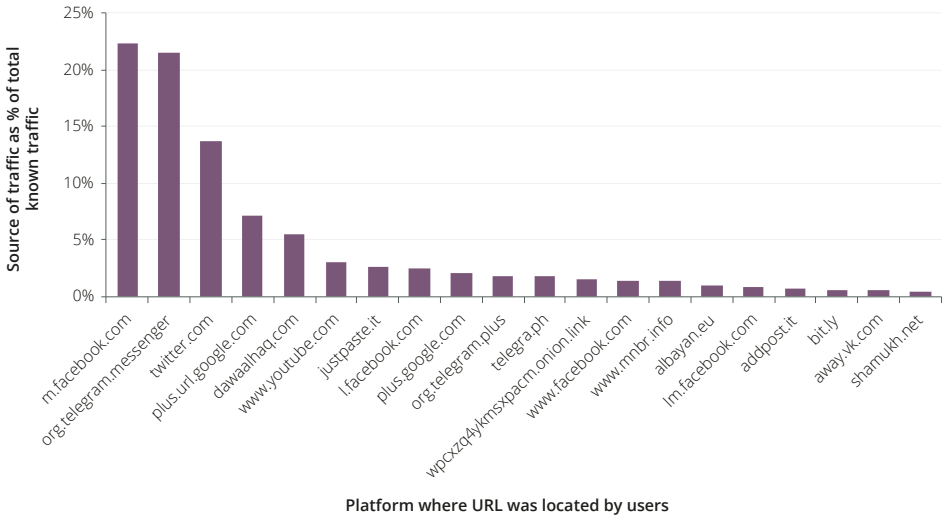
These platforms are frequently the locations where jihadist groups hide or store jihadist material, although they can also connect users to pages or groups on social media platforms.[16] These locations are useful as 'content stores' – places where sympathisers can access material because they take time to locate and remove. Some of these are websites crated directly by jihadist groups, which are obscure and protected by services such as Cloudflare, or jihadists use smaller platforms to host the actual material, as these are less likely to have systems to rapidly locate and remove material. Still, the utility comes from the ability to direct users to content, as without the URL it is often hard to find the specific file to download. As a result, for platforms to be useful to the dissemination of jihadist content, the intended audience has to be able to acquire the long URL or a convenient shortlink. Where the audiences find and click on those URL are the 'sources' of traffic.

15. Domain names are more familiar, mostly because they are easy to remember, but the correct IP address leads to the same location – for example, RUSI.org can also be reached via its IP address, 35.176.221.125.
16. Other sites jihadists find useful include academic repositories. Azelin.files.wordpress ranked as the 80th most frequent target of traffic and 67th in terms of number of shortlinks directing users to the site. Videopress (a video hosting platform formerly used by Jihadology) ranked 59th for traffic and 67th (equal with Azelin.files.wordpress) in number of shortlinks.

Sources of Traffic

Identifying these sources of traffic shows how jihadist groups are able to make their content available to users and potential supporters. Sources fulfil the function of beacons signposting users to the locations where they can access jihadist material, communicate with sympathisers and coordinate activities. Sources of traffic are outlined below and reveal that four mainstream platforms – Facebook, Telegram, Twitter and Google – are the four known top platforms used to direct users.[17]

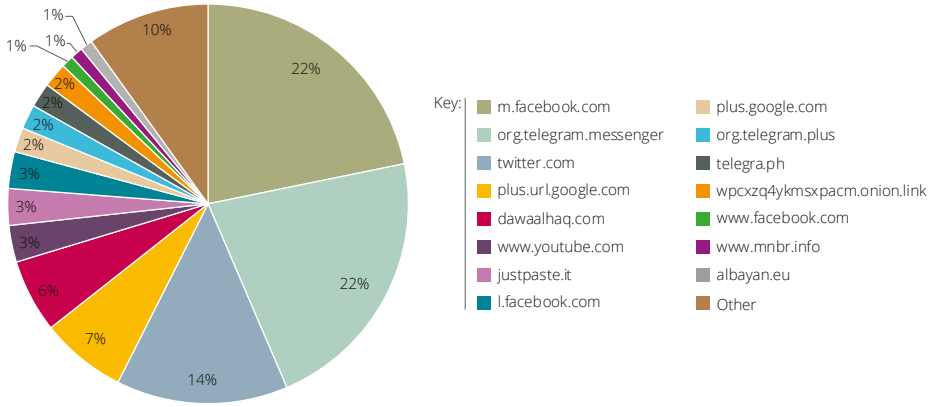
Figure 2: Sources of Traffic



It is important to note that not all sources of clicks are known as users sometimes mask their activity. Other users may find the link and paste it into their browser rather than clicking the link. Analysing the shortlink data therefore provides a model for where users are known to locate content rather than an analysis of every single click.

17. This research focuses on developing a model of the information ecosystem based on the available evidence, in this case the known sources of traffic. Known traffic excludes 'direct' or 'dark' traffic, where the service is unable to tell precisely where the individual acquired the shortlink or where the shortlink is pasted directly into a web browser. Email and SMS may also be included in this category depending on the service.

Figure 3: Source of Known Clicks



Over half of known clicks come from just three platforms (Facebook, Telegram and Twitter); these, together with simple content-sharing sites such as justpaste.it and telegra.ph, make up over 80% of the known sources of traffic. For example, the tweet shown in Figure 4 contains an audio segment of an Abu Bakr Al-Baghdadi speech, which when captured by BlackLight had been viewed 138,000 times.

Around 7% of known traffic comes from sites that are explicitly jihadist websites or theologically aligned sources from which jihadist groups draw inspiration. These include albayan.eu, mnbr.info, dawaalhaq.com and onion links.[18] The traffic data reminds us that, while this latter group of web pages tends to be thought of as places to access content, they also provide a way to access further locations in the network as well as means of content aggregation.

The short audio segment features Al-Baghdadi's speaking about Saudi Arabia, referred to in jihadist speak as 'the peninsula of Muhammad', addressing Sunnis as being under threat by the Saudi regime, who seek to alter Islam to dissuade Sunni Muslims understanding the Daesh version of Islam, while actively combating 'true' Sunnis in Yemen, Iraq and Syria.

18. Onion links are part of the Tor network that uses onion routing to allow individuals to anonymously browse the internet.

Figure 4: Audio Segment of an Abu Bakr Al-Baghdadi Speech on Twitter

Source: Abu Osama Al-Iraqi, 'Listen to what Sheikh Abu Bakr Al-Baghdadi says to the Al Salul and a message for the Muslims of the Arabian Peninsula' [Twitter post], 12:13am, 3 September 2018, account suspended and URL unavailable, accessed 3 September 2018.

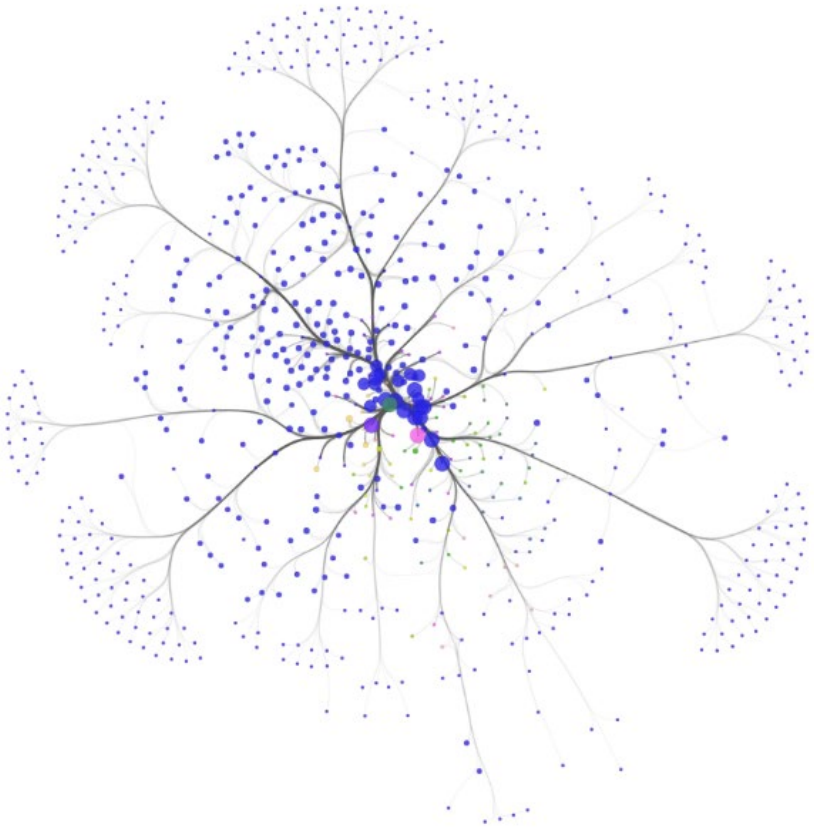
The Model

Combining the sources and targets of individual clicks demonstrates that traffic moved from, for example, the 'source' Facebook to the 'target' Sendvid.

The researchers aggregated traffic data and displayed known connections between the source of clicks and the target URL for each shortlink. Using SNA, this data was built into a network model of the jihadist information ecosystem. This method uses a line to represent the movement of traffic between the platform where a user clicks a link (the source) and the location to which they are redirected (the target).[19] The purpose of the model is to understand how the information ecosystem operates. This visualises information dissemination as an ecosystem rather than a piecemeal, 'whack-a-mole' approach at the level of individual accounts operating on a single platform.

19. This model is a representation rather than a complete model of traffic across the internet.

Figure 5: Model of the Jihadist Information Ecosystem



The model contains 723 nodes. Each node represents a platform or domain (including subdomains). Subdomains are included because services such as WordPress or Blogspot use these services to provide their users with distinct addresses (for example, klmtuhaq.files.wordpress.com or murtaqaa.blogspot.com).

The lines connecting the nodes represent the 'edges' in the network – the connection between each source and target node. There are 2,474 edges in the model, representing the 2,474 unique connections between the platforms jihadist groups use to direct users to the location of content and the sites where that content is stored.

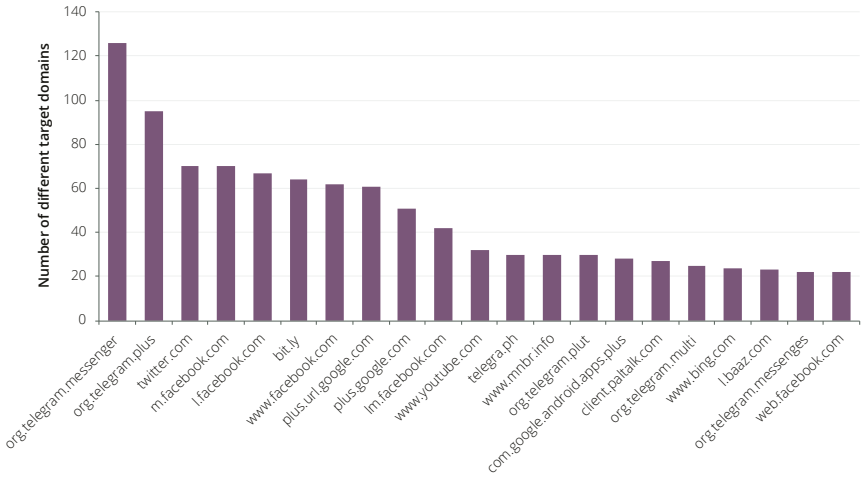
SNA produces a range of metrics which provide insight into the nature of the ecosystem. The first is that there is only one connected component – meaning that all nodes are connected in a single network.

Second, on average each node in the network is connected to 3.42 other nodes. This means that if one of the nodes connected to the platform was suspended, there would still be other available connections within the ecosystem.

Likewise, the network diameter is three, meaning that very few connections are required to span the network, and the average path length between any two nodes is 1.38 connections: most nodes are able to connect to others in one or two connections.

Significantly, however, this level of connectivity does not result from one or two important nodes (platforms), which would make the network vulnerable – in other words, if these nodes were removed, the network would no longer be able to function. Instead, as Figure 5 shows, the jihadist information ecosystem is a dispersed network in which there are many relatively important nodes.[20]

Figure 6: Sources to Diverse Targets

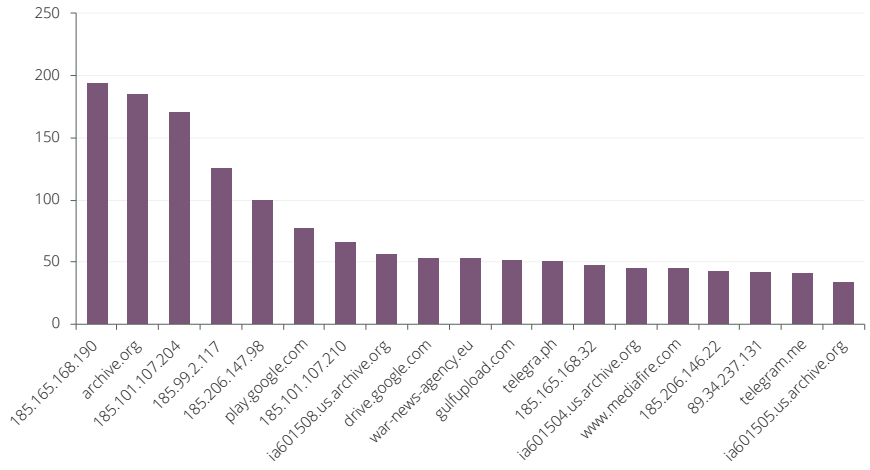


This makes the network less efficient because it requires some level of redundancy (the existence of more than one path across the network between two nodes).[21] However, the cost of that redundancy delivers a

20. This is discussed in Ali Fisher, 'Mapping the Great Beyond: Identifying Meaningful Networks in Public Diplomacy', CPD Perspectives on Public Diplomacy, Paper 2, 2010, pp. 1–87. Fisher's article draws on: Paul Baran, 'On Distributed Communications; Introduction to Distributed Communications Networks', RAND, 1964, <http://www.rand.org/pubs/research_memoranda/RM3420/>, accessed 10 July 2019; John Aquila and David Ronfeldt, 'The Advent of Netwar (Revisited)' in John Arquilla and David Ronfeldt (eds), *Networks and Netwars: The Future of Terror, Crime, and Militancy* (Santa Monica, CA: RAND, 2002), p. 1. See also Fisher, 'Netwar in Cyberia'.
21. For redundancy in a network, see Paul Baran, 'On Distributed Communications Networks', *IEEE Transactions on Communications Systems* (Vol. 12, No. 1, 1964), pp. 1–9.

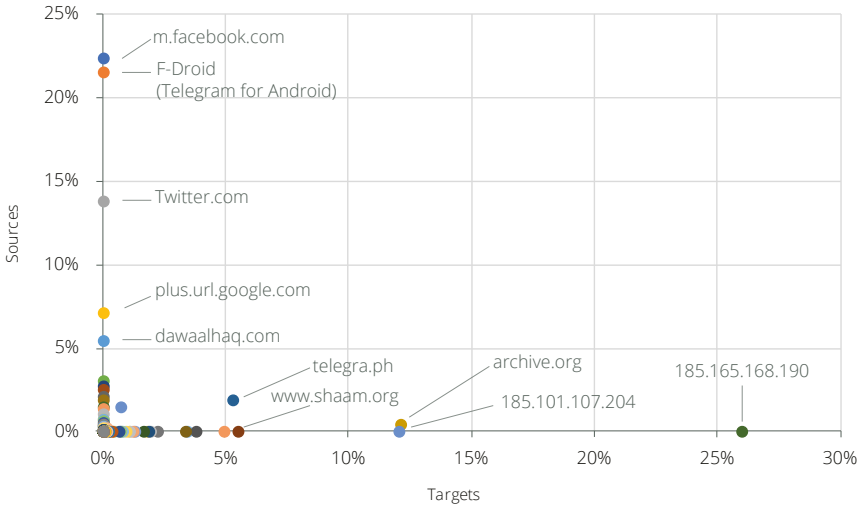
greater level of resilience, as the network's continuity is not contingent on one or two high-value nodes.

Figure 7: Traffic from Diverse Sources



The dispersed nature of the network is shown by the number of sources that have connections with many other nodes. Similarly, there are many targets which have incoming traffic from diverse sources.

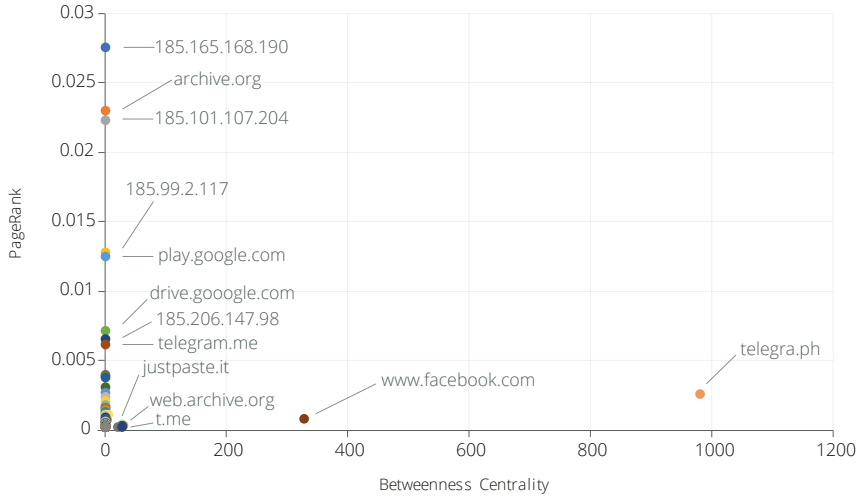
Figure 8: Sources Versus Targets (Known Traffic)



Seen as a scatterplot, the different functions that platforms and sites fulfil within the ecosystem become clearer. Sites located higher on the y-axis of Figure 8 are more likely to be functioning as signposts, while those further along the x-axis act as content stores or aggregators.

Other SNA metrics, specifically PageRank and betweenness centrality, provide further insight into what are known as 'key actors' (important platforms) in the network. Their position on the graph in Figure 9 gives a perspective on how they are important in the network.

Figure 9: Key Actors



Betweenness centrality refers to how often a node lies on the shortest path between any two nodes in the network. Actors who rank highly on betweenness centrality have the potential to influence others near them in a network.[22] A node with high betweenness centrality has greater potential influence over the spread of information through the network by facilitating, hindering or even transforming the form and content of communication between others.[23]

PageRank measures revolutionised the process of delivering web search results for Google and other search engines. PageRank refers to the probability distribution for nodes in a network. In other words, it is a measure of how likely a user is to reach a specific node from other

22. Noah E Friedkin, 'Theoretical Foundations for Centrality Measures', *American Journal of Sociology* (Vol. 96, No. 6, May 1991), pp. 1478–1504.
23. Alistair Willis, Ali Fisher and Ilia Lvov, 'Mapping Networks of Influence: Tracking Twitter Conversations Through Time and Space', *Participations: Journal of Audience & Reception Studies* (Vol. 12, No. 1, 2015), pp. 494–530.

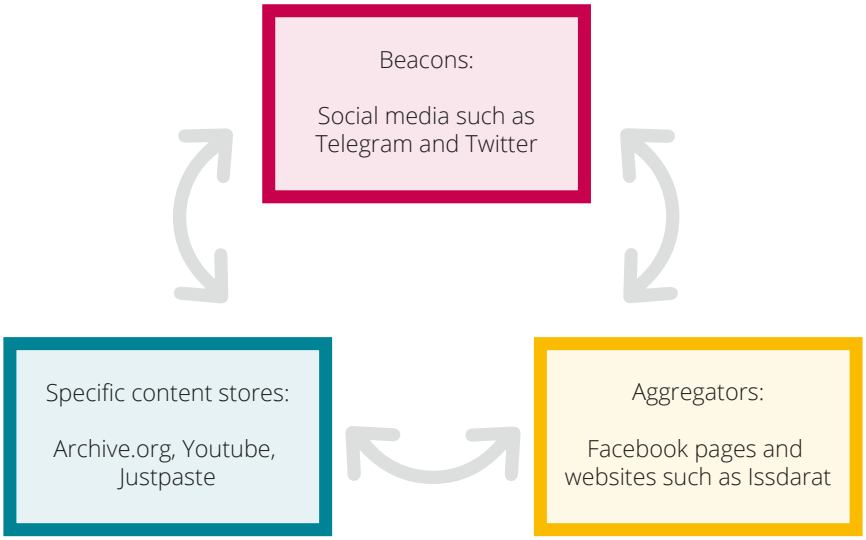
nodes in a network. As such, it is also a marker of influence.[24] These nodes are frequently found at the 'core' of the network.[25]

Nodes in the top left corner of Figure 9 tend to be near the core of an information-sharing network, while those at the lower right are bridges or conduits for information between various cores within a network.

As those platforms at the top left of the graph are also at the top of the list of platforms having diverse sources of traffic, this provides further evidence that jihadist groups leverage a multiplatform communication network to maintain a persistent presence for their material.

Analysis of the Research Findings

Figure 10: Platform Roles and Connectivity



While current policy has focused on individual pieces of content on individual platforms, albeit using sophisticated machine learning algorithms to identify extremist accounts and content, the 'swarmcast' used by jihadist groups exploits a multiplatform communication

24. For the authors' use of PageRank, the damping factor was set at (0.85/15%). This damping factor was selected as it was the level suggested by Brin and Page. It was also highlighted in Luca Becchetti and Carlos Castillo, 'The Distribution of PageRank Follows a Power-Law Only for Particular Values of the Damping Factor', paper presented to the 15th International Conference on the World Wide Web, Edinburgh, Scotland, 23–26 May 2006, where the authors observed that the 'typical damping factor used in practice is between 0.85 and 0.90'.
25. Azelin.files.wordpress.com is ranked 50th by page rank.

paradigm to leverage connectivity across over 700 platforms to distribute their content.

Until now, most research has focused on individual pieces of content or analysis of single platforms,[26] ignoring jihadists' ability to leverage connections across platforms.

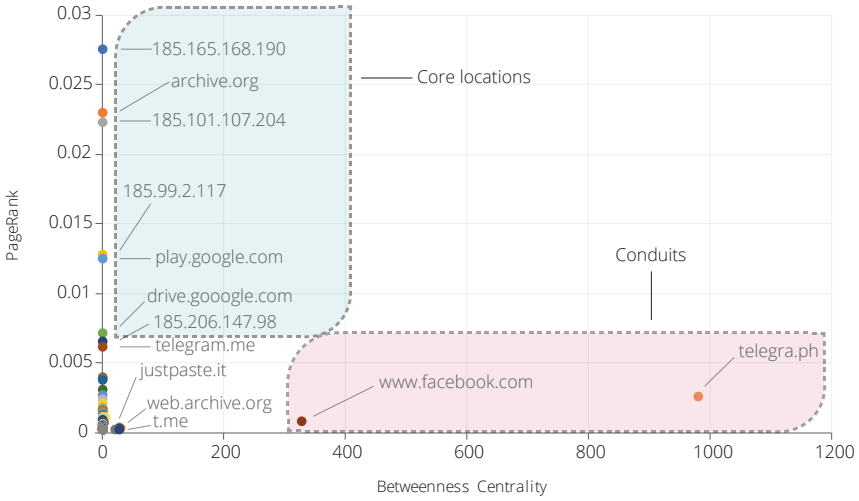
Unable to find content, some researchers conclude that jihadist material is not available or that options for accessing it are increasingly limited. This research, however, shows that the content is available on the surface web and social media platforms for those who know how to access it. For those who are less familiar with jihadist ideology, and who do not speak Arabic or have a strong understanding of theology, jihadist content on the surface web provides information about jihadist religious concepts. Non-Arabic materials are filled with transcribed Arabic key words, such as 'jihad', but also other related core concepts, such as 'shirk' or 'istishhad'. These project a specific interpretation of Sunni identity and encourage hostility towards all Sunnis who are not in line with this view, as well as Shiites and non-Muslims, rooting this in jihadist and extremist Salafist writings. The transcribed keywords allow the consumer to dig deeper into complex jihadist theology. By using a mix of Latin transcribed words and Arabic, the world of jihad online remains accessible.

Within the ecosystem, platforms and sites fulfil different functions. Comparing the metrics from social network analysis, Figure 11 shows the way domains are used by those accessing material important to the jihadist movement. Core locations are those which are particularly important to the network – indicated by the high PageRank value.

26. Jonathon M Berger and Jonathon Morgan, 'The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter', The Brookings Project on US Relations with the Islamic World, Analysis Paper 20, 2015; Maura Conway et al., 'Disrupting Daesh: Measuring Takedown of Online Terrorist Material and its Impacts', *Studies in Conflict & Terrorism* (Vol. 42, No. 1–2, 2019), pp. 141–60; Walid Magdy, Kareem Darwish and Ingmar Weber, '#FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support', arXiv:1503.02401, 2015; Elizabeth Bodine-Baron et al., *Examining ISIS Support and Opposition Networks on Twitter* (Santa Monica, CA: RAND Corporation, 2016); Matthew C Benigni, Kenneth Joseph and Kathleen M Carley, 'Online Extremism and the Communities that Sustain it: Detecting the ISIS Supporting Community on Twitter', *PloS One* (1 December 2017), doi:10.1371/journal.pone.0181405; Macdonald et al., 'A Study of Outlinks Contained in Tweets Mentioning "Rumiyah"'.

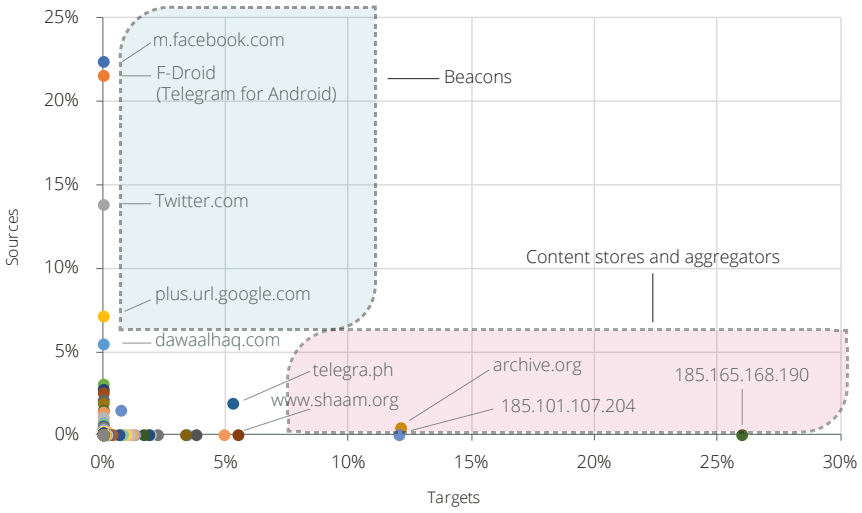
Conduits are those key actors that provide bridges, or act as conduits to jihadist material – indicated by high betweenness value.

Figure 11: Key Actors (Analysis)



Comparing the source and targets shows the functions that domains fulfil as part of content dissemination or content collection within the jihadist information ecosystem. This highlights that some sites are core to the content dissemination and others form conduits or bridges through which users may pass.

Figure 12: Sources Versus Targets (Known Traffic)



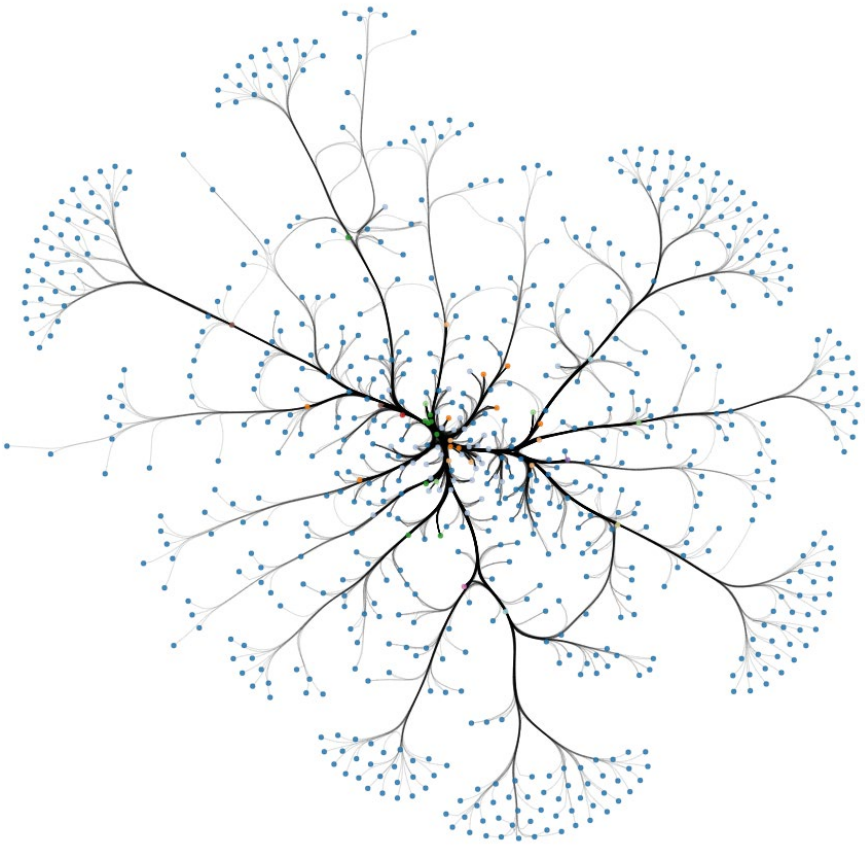
Based on the insights from the SNA and the comparison of domains which function as sources and targets, the research team classified these roles. Understanding how the jihadist movement exploits these

functions is the first step to developing the next generation of disruption approaches. While locating jihadist material on an individual platform is a challenge, the research shows that in a multiplatform communication network it is even more difficult to locate the sources and stores of content. Given the continuing evolution in digital media use, the following classification sets out the functions different platforms can fulfil in the jihadist information ecosystem, although it should be noted that these are not entirely distinct or mutually exclusive:

- **Beacons**: These provide the 'always on' stream of communication through which information can be rapidly disseminated. Beacons act as the sources of traffic, primarily providing a 'signpost' pointing users to locations they can access jihadist material.
- **Content stores**: Locations where content is uploaded for users to access with a link supplied by aggregators or beacons.
- **Content aggregators**: Sites that gather a range of jihadist materials and provide users with a collection of links to locations where a specific video can be downloaded. These archives of videos often take the form of a blog, using IP addresses, Latin character domains, or blog type sites, such as Wordpress, Blogger and Tumblr. For example, Ansar Al-Khilafah: ansarukhilafah.wordpress.com and amaqnews.tumblr.com.

Finally, the data also reveals that, unsurprisingly, Salafi–jihadist groups thrive on a common Arabic Salafist textual basis and produce their own Arabic language extremist content for their core target audience. To understand jihadist visual content, relevant Arabic-language texts must be understood. Moreover, jihadist texts often remain undetected due to the majority being in Arabic, but also because PDFs or Microsoft Word documents can be uploaded via more platforms than audiovisual content.

Figure 13: The Jihadist Online Ecosystem



Longevity of Content

Following feedback from GIFCT members, a subset of URLs was examined in May 2019 to assess whether they are still accessible. The links in the original study were tested to establish how many still produce a server response.[27] It was found that 658 links no longer produce any response. From the remaining links, the authors examined 380 URLs of the most easily identifiable jihadist content hosted on common content stores. This was done to make this follow-up study a manageable task within the timeframe. The content stores included: telegra.ph; archive.org; mediafire.com; justpaste.it; tlgur.com; gulfupload.com; and top4top.net. Of these, the majority (approximately 75%) were still available.

Conclusion

Focusing on the multiplatform communication paradigm rather than individual platforms is key to the future development of a

27. This includes the IP addresses – which are the vast majority of links which no longer respond.

next-generation approach to online disruption. The 'swarmcast' has evolved as jihadist groups have increasingly adopted a multiplatform approach for content dissemination. More specifically, it has changed in the following ways.

Diversity: As observed in nature and biology, diversity is a key component contributing to the resilience of an ecosystem. The jihadist information ecosystem is no different. Therefore, contrary to contemporary commentary, the diversity of platforms used ensures jihadist content is widely accessible via mainstream social media and the surface web. The major platforms and simple content-sharing sites are used as 'beacons' to share URLs to direct people to content. This makes up over 50% of the known sources of traffic to terrorist content. Only 7% of known traffic comes from sites where either explicitly jihadist content is hosted or content from which groups such as Daesh draw. However, at least 50% of the actual content is stored on specific websites run by jihadist or theologically aligned groups. As these are smaller and obscure servers this takes time to locate and remove. This means that current techniques employed by big platforms, such as image hashing, struggle to identify content shared in this manner since the content itself is stored on other, smaller or specifically created platforms. This makes the content undetectable by the original domain owner.

Multiplatform approach: Jihadist groups use a multiplatform approach to create and maintain a persistent online presence for their material. Individual platforms fulfil specific functions, creating a major obstacle to contemporary disruption activities. It is particularly important to differentiate between how links to material are shared and where the specific content is actually stored online. The ecosystem's resilience is also partially due to each platform or domain in the network connecting on average to at least three other nodes. This connectivity does not rely on one or two important nodes, which would make the network vulnerable, but remains dispersed, integrating many relatively important nodes. If one of these nodes was suspended, there would still be other available connections within the ecosystem.

Arabic language: Despite it being the primary language of jihadist groups, much contemporary analysis draws on social media accounts and sources in English. Failing to analyse Arabic text means that jihadist written content – which tends to be uploaded via more platforms than audio-visual content – risks being undetected. This type of content is often not removed due to the complexity of defining what is actually jihadi content as opposed to some extreme Salafist material. This type of content is at times endorsed, reshared and in some cases re-published with Daesh media organisation logos but remains available. Jihadists are aware that Arabic, in addition to considering it

a sacred language, provides a linguistic firewall which their adversaries find difficult to penetrate.

Recommendations

It is important to develop techniques to disrupt the distribution strategy of groups such as Daesh and the flow of potentially vulnerable users to illegal content. The following recommendations are proposed to support development of a next-generation approach to disruption.

- Technology companies and internet referral units should focus on multiplatform methods of content dissemination, in addition to the individual platform approaches such as content filters. This would mean embracing a multiplatform approach in relation to URL and shortlink reporting at greater scale and developing a robust shared awareness of URLs/shortlinks leading to jihadist content across the full range of platforms.
- This research shows the data feed is available, and that some companies have already successfully deployed this approach, based on services provided by organisations, such as Human Cognition. Many others could benefit if individual companies or the sector more broadly developed a mechanism to use these services effectively.
- Tech companies should also focus on the functions of each platform in the jihadist information ecosystem to disrupt the specific type of abuse that occurs on each.
- Similarly, researchers analysing jihadist activity on a single platform should design their research and frame their analysis based on that platform's function in the ecosystem.
- Shortlink service providers should make available data on jihadist use of their services to other technology companies and researchers, in order to support efforts to develop a more detailed perspective on the information ecosystem.

Platform owners should develop or acquire early warning systems for faster detection by penetrating digital networks where jihadist groups initially distribute new URLs/shortlinks and by aggressive tracking of users sharing jihadist content. For example, platform owners, once a URL is identified, could use this to locate where it appears in other social media messages posted on 'beacons' and 'aggregators' – as jihadist groups share content in batches of links.[28] Such systems would need to take account of current or proposed privacy legislation, and of duty of care responsibilities to employees being asked to take on potentially risky roles.

28. This has been the modus operandi since operating on classical visual basic/bulletin forums since the early 2000s.

Technology companies and internet referral units should recruit analysts with Arabic language skills, who are also well versed in both jihadist and extremist Salafist writings, to identify jihadist content. Where possible, technology companies should combine machine learning and automated removal with human analysis and verification, especially in relation to extremist (as opposed to terrorist) content.

Governments must provide clarity on the line between privacy protection and disruption of jihadist content dissemination. As jurisdictions propose new legislation, policymakers must ensure that their attempts to protect privacy and their desire to encourage platforms to take action against extremist content are not contradictory. Similarly, to maintain credibility, governments threatening to fine tech companies for failing to remove content quickly from their platforms cannot continue to work with researchers who have been actively posting jihadist content on the same platforms.

Ali Fisher is Explorer of Extreme Realms at Human Cognition.

Nico Prucha is Chief Content Curator at Human Cognition.

Emily Winterbotham is a Senior Research Fellow in the National Security Studies programme at RUSI.

This research was made possible with financial support from GIFCT and data donated by Human Cognition.

About RUSI

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 188 years.

About The Global Research Network on Terrorism and Technology

The Global Research Network on Terrorism and Technology is a consortium of academic institutions and think tanks that conducts research and shares views on online terrorist content; recruiting tactics terrorists use online; the ethics and laws surrounding terrorist content moderation; public–private partnerships to address the issue; and the resources tech companies need to adequately and responsibly remove terrorist content from their platforms.

Each publication is part of a series of papers released by the network on terrorism and technology. The research conducted by this network will seek to better understand radicalisation, recruitment and the myriad of ways terrorist entities use the digital space.

The network is led by the Royal United Services Institute (RUSI) in the UK and brings together partners from around the world, including the Brookings Institution (US), the International Centre for Counter-Terrorism (Netherlands), Swansea University (UK), the Observer Research Foundation (India), the International Institute for Counter-Terrorism (Israel), and the Institute for Policy Analysis of Conflict (Indonesia).

The research network is supported by the Global Internet Forum to Counter Terrorism (GIFCT). For more information about GIFCT, please visit https://gifct.org/.

The views expressed in this publication are those of the authors, and do not reflect the views of RUSI or any other institution.

Published in 2019 by the Royal United Services Institute for Defence and Security Studies.

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org

RUSI is a registered charity (No. 210639)