

COUNTER EXTREMISM PROJECT

THE eGLYPH WEB CRAWLER: ISIS CONTENT ON YOUTUBE

Introduction and Key Findings

From March 8 to June 8, 2018, the Counter Extremism Project (CEP) conducted a study to better understand how ISIS content is being uploaded to YouTube, how long it is staying online, and how many views these videos receive. To accomplish this, CEP conducted a limited search for a small set of just 229 previously-identified ISIS terror-related videos from among the trove of extremist material available on the platform.

CEP used two computer programs to locate these ISIS videos: a **web crawler** to search video titles and descriptions for keywords in videos uploaded to YouTube, and **eGLYPH**, a robust hashing content-identification system. CEP's search of a limited set of ISIS terror-related videos found that hundreds of ISIS videos are uploaded to YouTube every month, which in turn garner thousands of views.

Based on CEP's research parameters, we found that in this three-month period:

1,348 ISIS videos were uploaded to YouTube, garnering **163,391 views**

24% of those videos remained on YouTube for over two hours, receiving **148,590 views**

76% of those videos remained on YouTube for less than two hours, receiving **14,801 views**

278 accounts uploaded all **1,348 videos** to YouTube

60% of accounts remained live after uploaded videos had been removed for content violations

SECTION 1: BACKGROUND

A: Recent history

YouTube, Google's video streaming platform, has been an important site for posting and sharing ISIS's propaganda since the group's inception¹. Even as ISIS's message shifted from building a utopia in the Middle East to inspiring individuals to commit attacks in the West, YouTube has remained a central component of ISIS's online media strategy².

ISIS, like most Americans, is drawn to YouTube because it is the dominant online streaming platform. A March 2018 Pew Research Center report found that 73% of American adults use YouTube, with 94% of individuals between the ages of 18 and 24 using the site³.

Moreover, there is a clear link between extremist videos and individuals who have sought to support or join ISIS. A joint study from the University of Chicago's Project on Security and Threats and the Australian Strategic Policy Institute's Counter-Terrorism Policy Center found that 83% of Americans who committed or were charged with ISIS-related crimes between March 2014 and August 2016 watched ISIS propaganda videos⁴.



83% of Americans who committed or were charged with ISIS-related crimes between March 2014 and August 2016 watched ISIS propaganda videos⁴

B: Selected YouTube Policy Changes Over Time

YouTube’s policies restricting the types of videos prohibited on the site and how those videos are removed have only evolved in response to scandal or pressure from policymakers.

Below are several policy changes resulting from highly publicized discoveries of extremist content on YouTube and pressure from lawmakers and advertisers to remove this material.

Incident	Subsequent Policy Change(s)
<p>2016: Throughout the year, UK and European lawmakers expressed concern that social media platforms had become a “vehicle of choice” for extremists to recruit and radicalize. Several governments threatened legislative action⁵.</p>	<p>December 2016: YouTube, Facebook, Microsoft, and Twitter launched a shared industry database of “hashes”—digital “fingerprints” of extremist imagery—in an effort to curb the spread of terrorist content online⁶.</p>
<p>March 2017: A <i>Times of London</i> investigation found advertisements of reputable brands appearing alongside hateful and extremist videos⁷.</p>	<p>March 2017: YouTube announced that it will take a tougher stance on hate speech and strengthen advertiser controls⁸.</p> <p>June 2017: YouTube announced new guidelines about content eligible for ads⁹.</p>
<p>May-June 2017: In May, the <i>Times of London</i> found several bomb-making videos on Facebook and YouTube, days after Salman Abedi detonated a suicide bomb in Manchester, England. He reportedly built the bombs after watching instructional videos online. UK and European lawmakers also continued to increase pressure against tech companies, calling for new laws to punish companies that continue to host extremist material on their platforms.</p>	<p>June 2017: YouTube “increases its use of technology” to identify extremist videos, increases the number of people in its Trusted Flagger program, and takes a “tougher stance” on videos that do not clearly violate its policies¹¹.</p> <p>June 2017: YouTube, Facebook, Microsoft, and Twitter launched the Global Internet Forum to Counter Terrorism (GIFCT), a partnership aimed at combating extremist content online¹².</p>
<p>The UK Home Affairs Committee published a report saying that tech companies are “shamefully far” from taking action to tackle hateful content¹⁰.</p>	<p>July 2017: YouTube launched its Redirect Method program, which aims to redirect users searching for violent extremist content to counter-narrative videos¹³.</p>

C: YouTube's Polices and Removal of Extremist Content

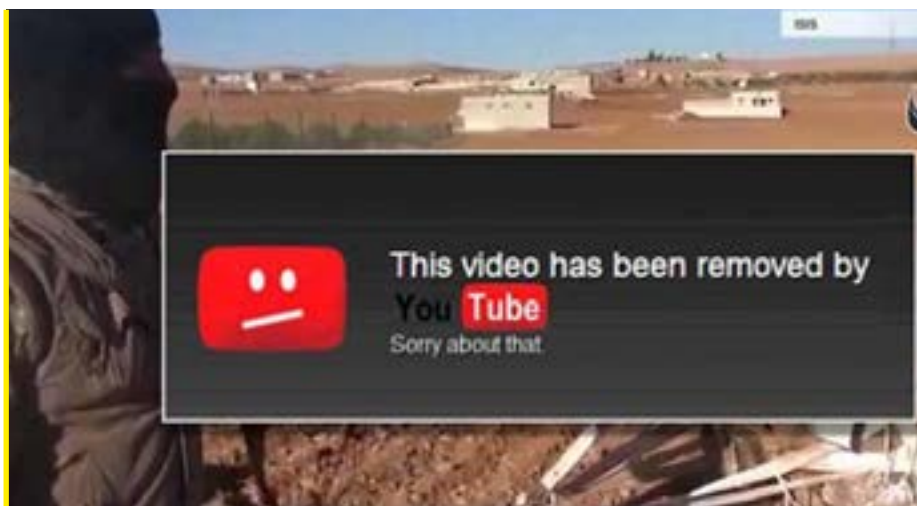
According to YouTube's Community Guidelines, it "do[es] not permit terrorist organizations to use YouTube for any purpose, including recruitment. YouTube also strictly prohibits content...that promotes terrorist acts, incites violence, or celebrates terrorist attacks¹⁴." Since June 2017, YouTube has also taken measures towards hiding extremist content that does not explicitly violate their rules, such as putting inflammatory videos behind warning labels.

YouTube claims to locate ISIS content through three mechanisms: human flagging, machine learning, and hashing¹⁵.

Human flagging involves an individual user notifying YouTube that a video violates the Community Guidelines¹⁶. The flagger may be a regular user browsing videos or a 'trusted flagger'¹⁷. Trusted flaggers are individuals, government agencies, or NGOs that YouTube has granted extra features for reporting content, such as mass flagging, YouTube staff support, and prioritizing those flagged reviews¹⁸. A YouTube employee, however, makes the final decision on whether or not to remove the flagged content.

YouTube states that it uses automated machine learning technology to locate extremist videos. Still, the decision to remove these videos is made by a human content moderator¹⁹. According to its 2018 Community Guidelines Transparency Report, YouTube located 72% (7,029,971 out of 9,790,083 videos) of removed extremist videos through automated machine learning systems between January and March 2018²⁰.

YouTube also claims to locate videos through hashes, which are signifiers unique to each piece of content²¹. The extent to which YouTube utilizes this technology is unclear, and it is uncertain whether this technology is used to prevent the reupload of known extremist content. In December 2016, Google/YouTube, along with Facebook, Twitter, and Microsoft, launched a shared hash database, which they claim now numbers over 88,000 hashes as of June 2018²².



METHODOLOGY

In order to test YouTube’s stated ability to remove ISIS content, CEP used a web crawler to search video titles and descriptions for pro-ISIS keywords in videos uploaded to YouTube, and eGLYPH, a hashing image identification system to locate known ISIS videos.

eGLYPH, developed by CEP Senior Advisor and digital forensics expert Dr Hany Farid, is an image identification system that uses robust hashing (see Annex A). A hash represents a distinct signature (like a fingerprint) of a multimedia file, and can be used to detect duplicate image, video, and audio recording files.

As a robust hashing system, eGLYPH can also detect content that has been slightly modified, such as by changing the video’s resolution and/or speed (i.e., slow down or speed up), as well as video segments typically used in pro-ISIS video montages. CEP created a hash database composed of 229 full-length ISIS videos, introductory clips from ISIS videos with variations based on ISIS’s administrative regions or media outlets, and important excerpts from ISIS videos that have appeared in multiple pieces of visual propaganda. CEP also identified 183 keywords that are indicative of pro-ISIS sympathies and are typically associated with ISIS content such as slogans as well as the names of ISIS’s provinces, media outlets, and prominent propagandists (see Annex B).

The web crawler, attached with eGLYPH, conducted searches using YouTube’s API every 20 minutes for 24-hours a day for video uploads that match one or more of the 229 hashes, as well as one or more of the 183 keywords. Video matches were downloaded to CEP’s server, which recorded key data, including: date and time of upload, the number of views, the account that uploaded the video, and the time of removalⁱ (if at all).



i: Due to YouTube’s API’s quota requirements, CEP checked for the availability of an ISIS video over a three-day period. CEP did not confirm whether or not YouTube removed the ISIS video after that three-day period.

DATA OVERVIEW

Collected Data on YouTube

Between March 8 and June 8, 2018, using a narrow set of research parameters, CEP located 1,348 ISIS videos on YouTube that included content matching ISIS video content included in CEP's database. Those 1,348 ISIS videos received a total of 163,391 views before they were removed (if at all) by YouTube.

Of those 1,348 videos, 322 videos (24%) were on YouTube for two hours or longer, accumulating a total of 148,590ⁱⁱ views (91% of all views), with an average of 461 views per video. And 1,026 videos (76%) were on YouTube for less than two hours, receiving 14,801 views (9% of total views), with an average of 14 views per video.

91% of all uploaded videos were uploaded more than once, meaning that YouTube's hashing systems are not working appropriately to prevent the reuploads of known terrorist videos.

All 1,348 ISIS videos were uploaded by 278 different YouTube accounts. 92% of accounts uploaded more than one video (with one account uploading 50 videos). 60% of accounts remained live on YouTube after videos were deleted, meaning that the account was not deleted after an ISIS video was removed by YouTube staff.

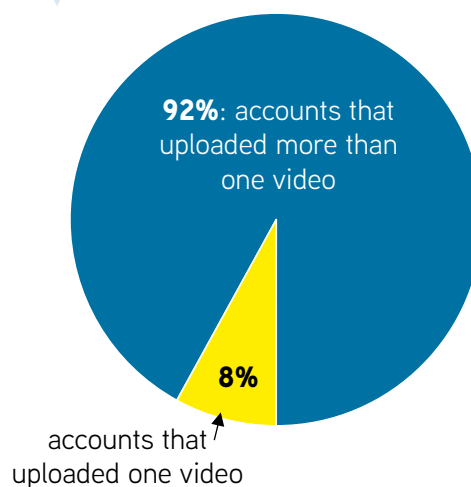
CEP was surprised to see that these accounts remained on the platform. It is unclear why YouTube would not delete them immediately, especially when the accounts are known to have uploaded terrorist material.

ii: CEP only checked for the availability of an ISIS video over a three-day period. As a result, the actual number of views for these videos on YouTube that remained online after the three-day period is higher.

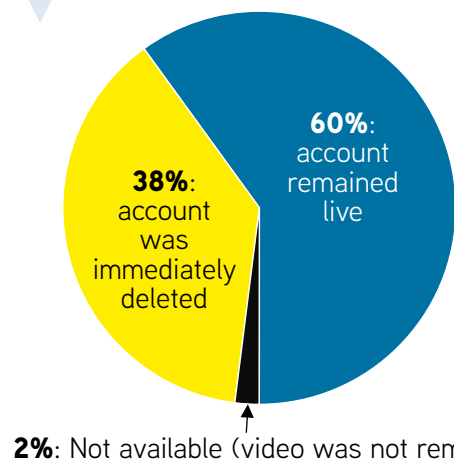
Percentage of total views



Percentage of accounts that have uploaded more than one video



Percentage of accounts that remained live after uploaded videos had been removed for content violations



CONCLUSION

CEP's findings call into question YouTube's claims of proactive content removal efforts. The fact that 91% of extremist videos had been reuploaded to YouTube at least once casts doubt on YouTube's stated efforts to prevent the upload or removal of known terrorist material.

YouTube's human flagging efforts are currently inadequate to consistently locate and remove known terrorist content. The fact that 24% of content remained on YouTube for more than two hours indicates that YouTube is failing to prepare, train, and educate its content moderators about known ISIS videos.

Similarly, while YouTube's machine learning technology appears to have made improvements in locating ISIS content, it does not excuse the fact that these videos are still allowed to be uploaded and accrue in some cases hundreds or thousands of views. The overall total of 163,391 views within a three-month period on this limited set of 1,348 uploaded videos shows that YouTube is still an important site for ISIS's propaganda efforts.

It is unlikely that Google/YouTube is deploying hashing technology appropriately given that known ISIS videos continue to be reuploaded to the platform. If video hashes were checked against a hash database of known terrorist content at the point of upload, it should prevent that video from being posted to YouTube.

Instead, CEP found that 91% of the narrow set of 1,348 videos had been uploaded more than once during the three-month period. For example, the ISIS video, Hunt Them O, Monotheist, originally released by ISIS's Somalian affiliate on December 25, 2017, encourages firearm and vehicular attacks in Western countries. During CEP's research period, the video was uploaded to YouTube on March 10, 2018 and was available for 29 hours and received 405 views before it was removed. The video was reuploaded on March 11 by a different account and was available for 39 hours and received 113 views before it was again removed. The video was then again reuploaded and deleted at least ten more times, receiving another 990 views, before the conclusion of CEP's research period on June 8.

Clearly, it remains possible to reupload known ISIS content despite YouTube's highly publicized promises to be proactive in removing content and use of machine learning and hashing technology²⁴. The fact that known terrorist videos continue to be uploaded and reuploaded to the platform calls into question YouTube's true intentions behind its heavily promoted efforts to combat online extremism. YouTube still has a long way to go in the fight against terrorist propaganda on their platforms.

The fact that known terrorist videos continue to be uploaded and reuploaded calls into question YouTube's true intentions behind its heavily promoted efforts to combat online extremism

RECOMMENDATIONS

CEP proposes several recommendations in order to help prevent the upload and spread of ISIS videos on YouTube.

Firstly, Google/YouTube should provide clear policy guidelines and take action to consistently and immediately delete accounts that have uploaded ISIS videos. Users should not be allowed to maintain their YouTube account after posting terrorist content.

Secondly, Google/YouTube must be more transparent regarding its hashing efforts. Google/YouTube should fully explain how it is implementing hashing technology, specifically if YouTube is deploying at the point of upload. Google/YouTube should also provide a detailed explanation of how the company contributes to and participates in the so-called “hashing coalition” announced in December 2016.

Google/YouTube should state how much content they have contributed to this shared database, and whether there is an agreement that all content in the database be removed across industry platforms and websites that are members of the hashing coalition and the Global Internet Forum to Counter Terrorism (GIFCT). Google/YouTube should also state how much content has been removed from their platform as a result of the database, and how the company as well as their coalition partners keep the database up to date.

An objective of the GIFCT is to share knowledge, information, and best practices. As a founding member of the GIFCT, Google/YouTube should aim to set industry standards on hashing practices.

A mandate for all GIFCT members to hash and remove content produced by groups and individuals sanction-designated by the United States, European Union, and United Nations, as well as material that glorifies or incites violence would rationalize

content removal practices and dramatically reduce the amount of terrorist content online.

Lastly, while removing content quickly from YouTube is clearly an important component of any effort to restrict the dissemination of terrorist propaganda, ‘time online’ should not be the only metric used to gauge the progress of YouTube and other platforms in ending their facilitation of terrorist propaganda. CEP has found that in many cases, ISIS videos removed within two hours still received dozens and, in some cases, hundreds of views. With CEP’s narrow scope of research parameters, we found that 1,026 videos (76%) were on YouTube for less than two hours, receiving 14,801 views (9% of total views).

Policymakers in the United States, United Kingdom, Germany, and elsewhere in Europe are in the process of developing regulations to compel tech companies to remove terrorist content quickly. Timely removal of ISIS videos is indeed important, but lawmakers should not lose sight of the importance of views. ISIS material, similar to other types of propaganda, is posted in order to be viewed and influence opinions and actions. A larger audience raises the possibility that an additional viewer may commit an act of terrorism in the name of ISIS. For instance, a video calling for terrorist attacks during holiday celebrations that has been on YouTube for one hour with 100 views has the potential to be more damaging than if that same video was online for two hours but only had a dozen views.

Setting standards for removal time periods is necessary, but lawmakers must also consider regulating and potentially fining companies based on highly-viewed terrorist material.

Lawmakers must consider regulating and potentially fining companies based on highly-viewed terrorist material

FOOTNOTES & ANNEXES 1&2

FOOTNOTES

- 1: Haroon Siddique, "Jihadi recruitment video for Islamist terror group Isis features three Britons," Guardian (London), June 20, 2014, <https://www.theguardian.com/world/2014/jun/20/jihadi-recruitment-video-islamist-terror-group-isis-features-britons>.
- 2: Tom O'Connor, "London Terror Attack Prompts Hundreds of New ISIS Videos on YouTube," Newsweek, March 29, 2017, <http://www.newsweek.com/london-terror-attack-inspire-isis-videos-youtube-576196>.
- 3: Aaron Smith and Monica Anderson, "Social Media Use in 2018," Pew Research Center, March 1, 2018, http://assets.pewresearch.org/wp-content/uploads/sites/14/2018/03/01105133/PI_2018.03.01_Social-Media_FINAL.pdf.
- 4: Robert Pape, Jean Decety, Keven Ruby, et al., "The American Face of ISIS," Australian Strategic Policy Institute, February 2017, https://s3-ap-southeast-2.amazonaws.com/ad-aspi/import/ASPI_CPOST_ISIS_Indictees.pdf?2Tbn8TshXmujb1ft8f7PIR7sukzyr hka.
- 5: Steven Swinford, "Facebook, Twitter and Google have become a 'recruiting platform for terrorism,'" Telegraph (London), August 25, 2016, <https://www.telegraph.co.uk/news/2016/08/24/facebook-twitter-and-google-have-become-a-recruiting-platform-fo/>; Libby Plumber, "Facebook, Twitter and YouTube promise to clamp down on hate speech," Wired, June 1, 2016, <http://www.wired.co.uk/article/facebook-youtube-twitter-and-microsoft-must-remove-hate-speech-in-24-hours-says-eu>.
- 6: "Partnering to help curb the spread of terrorist content online," Google Blog, December 5, 2016, <https://blog.google/topics/google-europe/partnering-help-curb-spread-terrorist-content-online/>.
- 7: Alexi Mostrous, "YouTube hate preachers share screens with household names," Times (London), March 17, 2017, <https://www.thetimes.co.uk/article/youtube-hate-preachers-share-screens-with-household-names-kdmpmkkjk>.
- 8: "Strengthening YouTube for advertisers and creators," YouTube Creator Blog, March 20, 2017, <https://youtube-creators.googleblog.com/2017/03/strengthening-youtube-for-advertisers.html>.
- 9: "Your Content and Making Money from Advertising on YouTube," YouTube Creator Blog, June 1, 2017, <https://youtube-creators.googleblog.com/2017/06/your-content-and-making-money-from.html>.
- 10: Tom Powell, "Facebook and YouTube under fire for carrying 'bomb making' guides," Evening Standard (London), May 25, 2017, <https://www.standard.co.uk/news/uk/facebook-and-youtube-under-fire-for-carrying-bombmaking-guides-a3548741.html>; Alexi Mostrous, "YouTube and Facebook awash with slick bomb making films," Times (London), May 25, 2017, <https://www.thetimes.co.uk/article/youtube-and-facebook-awash-with-slick-bomb-making-films-sn9prxpv6>; Fiona Hamilton and Alexi Mostrous, "Manchester Arena killer Salman Abedi used YouTube to build bomb," Times (London), June 24, 2017, <https://www.thetimes.co.uk/article/ariana-grande-manchester-concert-killer-salman-abedi-used-youtube-to-build-bomb-qzcb55s3>; John Shinal, "Facebook takes steps to slow spread of extremist content after EU leaders threaten new fines," CNBC, June 15, 2017, <https://www.cnb.com/2017/06/15/facebook-to-aid-terror-fight-after-france-u-k-signal-new-laws.html>; Anushka Asthana, "Theresa May calls on tech firms to lead fight against online extremism," Guardian (London), May 25, 2017, <https://www.theguardian.com/politics/2017/may/25/theresa-may-calls-on-tech-giants-to-lead-fight-against-online-extremism>; "Hate crime: abuse, hate and extremism online," United Kingdom House of Commons Home Affairs Committee, May 1, 2017, <https://publications.parliament.uk/pa/cm201617/cmselect/cmhaff/609/609.pdf>.

- 11: Kent Walker, "Four steps we're taking today to fight terrorism online," Google Blog, June 18, 2017, <https://www.blog.google/topics/google-europe/four-steps-were-taking-today-fight-online-terror/>.
- 12: "Facebook, Microsoft, Twitter and YouTube Announce Formation of the Global Internet Forum to Counter Terrorism," YouTube Official Blog, June 26, 2017, <https://youtube.googleblog.com/2017/06/facebook-microsoft-twitter-and-youtube.html>.
- 13: "Bringing new Redirect Method features to YouTube," YouTube Official Blog, July 20, 2017, <https://youtube.googleblog.com/2017/07/bringing-new-redirect-method-features.html>.
- 14: "Violent or graphic content policies," YouTube Help, <https://support.google.com/youtube/answer/2802008?hl=en>.
- 15: "YouTube Community Guidelines enforcement," Google YouTube Transparency Report, 2018, <https://transparencyreport.google.com/youtube-policy/overview>.
- 16: "Policies and Safety," YouTube, 2018, <https://www.youtube.com/yt/about/policies/#community-guidelines>.
- 17: "YouTube Community Guidelines enforcement," Google YouTube Transparency Report, 2018, <https://transparencyreport.google.com/youtube-policy/overview>.
- 18: "YouTube Trusted Flagger program," YouTube Help, <https://support.google.com/youtube/answer/7554338?hl=en>.
- 19: "YouTube Community Guidelines enforcement," Google YouTube Transparency Report, 2018, <https://transparencyreport.google.com/youtube-policy/overview>.
- 20: "YouTube Community Guidelines enforcement," Google YouTube Transparency Report, 2018, <https://transparencyreport.google.com/youtube-policy/overview>.
- 21: Scott Shane, "In 'Watershed Moment,' YouTube Blocks Extremist Cleric's Message," New York Times, November 12, 2017, <https://www.nytimes.com/2017/11/12/us/politics/youtube-terrorism-anwar-al-awlaki.html>.
- 22: "Tech firms hail 'progress' on blocking terror," BBC, June 8, 2018, <https://www.bbc.com/news/technology-44408463>.
- 23: In October 2017, Google, Twitter, Microsoft, and Facebook agreed with the G7 on the goal of removing extremist content from their platforms in two hours. "Tech firms to remove extremist posts within hours," BBC, October 20, 2017, <http://www.bbc.com/news/technology-41693777>.
- 23: 24 "Facebook, Microsoft, Twitter and YouTube Announce Formation of the Global Internet Forum to Counter Terrorism," YouTube Official Blog, June 26, 2017, <https://youtube.googleblog.com/2017/06/facebook-microsoft-twitter-and-youtube.html>.

ANNEX A

A Technical Explanation of eGLYPH by Dr Hany Farid



Data hashing has long been employed to identify or verify data. Specifically, a hash function is applied to data of any format, converting the data into an integer (referred to as a hash value, hash sum, or signature).

This hash function should have several properties: (1) the function should be computationally efficient to evaluate; (2) the function is deterministic, returning exactly the same signature when presented with the same input data; (3) the function should return a unique signature—that is, there should be no two distinct input data that hash to the same signature; and (4) the signature should be relatively compact.

There are many applications of hashing including efficiently accessing data in large databases, finding duplicate data entries in a database, verifying the integrity of a binary executable, and authenticating messages. For these applications, the uniqueness (or near uniqueness) of the hash function is critical. That is, data-searching and verification relies on the basic principle that any modification to the data will yield a different signature.

Hashing has been proposed as a means to detect duplicate images in a database. This approach works well if the image being searched is precisely the same image as that in the database. Any modification to the image, however, renders this approach relatively useless, as even the most minor of modifications (changing one pixel, adjusting contrast, resizing, re-compressing, etc.) changes the resulting hash signature.

Images (and audio/video) are somewhat distinct from other data formats in that it is often desirable to relax the definition of “duplicate” to be the same image invariant to simple modifications such as re-sizing, compression quality, and contrast/color enhancement. To this end, the term robust hashing has emerged to mean that the extracted signature of a data file is invariant to pre-defined transformations.

We have developed a robust hashing algorithm - eGLYPH - for identifying images, videos, and audio recording. The image-based version of eGLYPH operates as follows: (1) a full-resolution color image is converted to grayscale and down-sized to a lower and fixed resolution of 400 x 400 pixels.

This step reduces the processing complexity in subsequent steps, makes the robust hash invariant to image resolution, and eliminates high-frequency differences that may result from compression



artifacts; (2) perform a high-pass filter on the reduced resolution image to highlight the most informative parts of the image; (3) partition the image into non-overlapping quadrants from which basic statistical measurements of the underlying content are extracted and packed into a feature vector; and (4) compute the similarity of two hashes as the Euclidean distance between two feature vectors - distances below a specified threshold qualify as a match.

Despite its simplicity, this robust hashing algorithm has proved to be highly accurate and computationally efficient to compute (and is similar to the PhotoDNA that I co-developed in the mid-2000s and which today is used to find and remove child exploitation material on many on-line platforms).

The largest challenge with extending the image-based hashing to video is the massive amount of data in an even short video; at 24 frames per second, a three-minute video contains 4,320 still images. At even a modest resolution of 640 x 480 pixels per frame, a three-minute video contains over 1.3 billion pixels. The complexity of hashing a video, as compared to analyzing a single image, is at least three orders of magnitude larger.

There are, however, typically only small changes between successive frames of a video leading to a large amount of information redundancy in a video. We can, therefore, reduce the complexity of analyzing a video by first reducing this redundancy.



We, conveniently, just described a mechanism for measuring the similarity between two images. In addition to finding nearly identical images, robust hashing can be used to find similar images by controlling the threshold on the Euclidean metric for image similarity (as described in the previous section).

We start a video analysis by using robust image hashing to eliminate redundant video frames. This elimination of redundant frames typically reduces the length of a video by approximately 75%.

The image hash is then extracted from each of the remaining frames and concatenated to yield a final video hash. Unlike the image-based hashing that yields a fixed length hash, a video hash can be of arbitrary length. This presents both a challenge and an opportunity for comparing two hashes.

A Euclidean distance cannot, of course, be used to compare two hashes of arbitrary length. Instead, we utilize the longest common substring (LCS). By way of intuition, the longest common substring of the two strings “ABABACABBC” and “ABACABACBBCA” is six because the longest common string shared by these strings is the “ABACAB.” Note that these strings also have the substring “BBC” in common but this is shorter than the substring of length six.

The advantage of using LCS to compare two hashes is that it allows us to find not just matching videos but also video segments that are extracted or video segments that are embedded within a larger video (e.g., a video compilation).

Running on a standard Linux machine, a Java-based implementation of this robust video hashing requires approximately 10ms to process a single video frame and approximately 2.5ms to compare two hashes. To improve the efficiency, we have implemented a multi-core version of this algorithm that allows for a video to be partitioned into an arbitrary number of short segments each of which can be analyzed on a separate computer core.

The individual results from each segment are then combined to create a single hash. With this approach, the rate-limiting step to analyze any video is simply the number of computing cores that are available. •



ANNEX B

Keywords Indicating Pro-ISIS Sympathies and Associated with ISIS Content Used in Web Crawler

The Islamic State (Arabic)	Las Vegas Conquest 1 (Arabic)
Abu Bakr al-Baghdadi	Las Vegas Conquest 2 (Arabic)
Remaining and Expanding (Arabic)	Las Vegas Equitable Revenge (Just Terror Tactics) (Arabic)
Wikalat A'maq (Arabic)	Sufficient is your Lord as a Guide and Helper (Arabic)
Caravan of Martyrs (Arabic)	
Flames of War (Arabic)	United Cyber Caliphate (Arabic)
Deterring the Hired (Arabic)	Caliphate Cyber Army (Arabic)
Martyrdom Seeker (Arabic)	Mujahid (Arabic)
Loyalty and Disavowal (Arabic)	Democracy and Islam
The Polytheists (Arabic)	Inside the Khalifah (Arabic)
Fire-breathers (Arabic)	Hayat Al Jihad (Arabic)
Crusaders (Arabic)	For the Sake of Allah (Arabic)
Nusayri (Arabic)	The Making of Illusion (Arabic)
Abode of Disbelief (Arabic)	Paris Has Collapsed (Arabic)
Abode of Islam (Arabic)	You Must Fight Them O Muwahhid (Arabic)
Rejectionists (Arabic)	Just Terror (Arabic)
Cubs of the Caliphate (Arabic)	Fertile Nation (Arabic)
The Tyrants (Arabic)	Amir al-Mu'minin
The Jolani Front (Arabic)	Anwar al-Awlaki
Monotheism (Arabic)	Anwar Awlaki
The Awakening (Arabic)	Abu Hassan al-Mujahir
The Apostates (Arabic)	Abu Haleema
Hijra- Migration (Arabic)	Ayman al-Zawahiri
The Parties of Satan	Hamza bin Laden

Ahmad Musa Jibril	Wilayah Khurasan
Abu Baraa	Wilayah Gharb Afrīqiyyah
Musa Cerantonio	Wilayah Gharb Afrīqiyyah (Arabic)
Abdullah Faisal	Wilayat Gharb Ifrīqiyyah
Tarik Chadlioui	Wilayah Adan-Abyan
Tarik Ibn Ali	Wilayah Adan-Abyan (Arabic)
Abu Usamah al Gharib	Wilayat Adan-Abyan
Fursan (Arabic)	Wilayah al-Ḥijāz
Amaq (Arabic)	Wilayah al-Ḥijāz (Arabic)
Inside (Arabic)	Wilayah al-Barakah (Arabic)
Caliphate (Arabic)	Wilayah al-Barakah
Islamic Country (Arabic)	Wilayah al-Bayda
Jihad (Arabic)	Wilayah al-Bayda (Arabic)
Soon God Willing (Arabic)	Wilayah al-Badiyah
War and Media (Arabic)	Wilayah al-Badiyah (Arabic)
Agency Video (Arabic)	Wilayah al-Fallujah
Wilayah al-Barqah (Arabic)	Wilayah al-Fallujah (Arabic)
Wilayah al-Barqah	Furat (Arabic)
Wilayah al-Tarabulus	Furat
Wilayah al-Tarabulus (Arabic)	Wilayah al-Janub
Wilayah al-Fizan (Arabic)	Wilayah al-Janub (Arabic)
Wilayah al-Fizan	Wilayah al-Janoub
Wilayah al-Jazair	Wilayat al-Janoub
Wilayah al-Jazair (Arabic)	Wilayah al-Khayr
Wilayah Sinai	Wilayah al-Khayr (Arabic)
Wilayah Sinai (Arabic)	Wilayah al-Qawqaz (Arabic)
Wilayah Khorasan	Wilayah al-Qawqaz
Wilayah Khorasan (Arabic)	Wilayah al-Kavkaz

Wilayah al-Kavkaz (Arabic)	Wilayah Diyala (Arabic)
Wilayah al-Raqqa	Wilayah Kashmir
Wilayah al-Raqqa (Arabic)	Wilayah Kashmir (Arabic)
Wilayah al-Shishan	Wilayah Kirkuk
Wilayah al-Shishan (Arabic)	Wilayah Kirkuk (Arabic)
Wilayah al-Anbar	Wilayah Najd
Wilayah al-Anbar (Arabic)	Wilayah Najd (Arabic)
Wilayah Hadramawt	Wilayah Ninawa
Wilayah Hadramawt (Arabic)	Wilayah Ninawa (Arabic)
Wilayah Halab	Wilayah Shabwah
Wilayah Halab (Arabic)	Wilayah Shabwah (Arabic)
Wilayah Hamah	Alwaad
Wilayah Hamah (Arabic)	Alwaad (Arabic)
Wilayah Homs	Al Rased Al Felasatiny (Arabic)
Wilayah Homs (Arabic)	Al Hayat (Arabic)
Wilayah Salah al-Din	Jaysh Khalid Ibn al Waleed (Arabic)
Wilayah Salah al-Din (Arabic)	The Polytheists (Polytheismus)
Wilayah Salaheddine	The Polytheists (Vielgötterei)
Wilayat Salaheddine	The Polytheists (Polytheïsme)
Wilayah Sana'a	The Polytheists (I Politeisti)
Wilayah Sana'a	The Polytheists (Il politeismo)
Wilayah Baghdad	The Polytheists (Les polytheists)
Wilayah Baghdad (Arabic)	Abode of Disbelief (Heimat des Unglaubens)
Wilayah Dijlah	Abode of Disbelief (Huis van ongelooft)
Wilayah Dijlah (Arabic)	Abode of Disbelief (Dar Al-Kufr)
Wilayah Dimashq	Abode of Disbelief (Demeure de mécréance)
Wilayah Dimashq (Arabic)	Abode of Disbelief (darul kufur)
Wilayah Diyala	Abode of Disbelief (nevjernička država)

Inside the Khalifah (À l'intérieur du Califat)	Abode of Islam (Demeure de l'Islam)
Inside the Khalifah (unutar kalifate)	Abode of Islam (Darul islam)
Islamic State (Islamischer Staat)	Wikalat A'maq (Amaq Nachrichtenagentur)
Islamic State (Islamitische staat)	Wikalat A'maq (L'Amaq)
Islamic State (Stato Islamico)	Wikalat A'maq (Novinska agencija Amak)
Islamic State (L'État islamique)	Flames of War (Flammen des Krieges)
Islamic State (Islamska Država)	Flames of War (Vlammen van oorlog)
Abode of Islam (Heimat des Islam)	Flames of War (fiamme di Guerra)
Abode of Islam (Huis van Islam)	Flames of War (les flammes de la guerre)
Abode of Islam (Dar Al-Islam)	Flames of War (nevjernička država)
You Must Fight Them O Muwahhid (Du musst sie Bekämpfen Muwahhid)	
You Must Fight Them O Muwahhid (Je moet tegen hen vechten Muwahhid)	
You Must Fight Them O Muwahhid (Vous devez les combattre Muwahhid)	
You Must Fight Them O Muwahhid (borite se protiv njih o vjernici)	