Ofcom
making communications work
for everyone

The Buffalo Attack: Implications for Online Safety

A survey of selected industry responses and the state of cross-industry collaboration related to livestreamed terrorist attacks

Publication date: 12 October 2022

Contents

Section

Executive Summary 1

Introduction 4

Background to the attack 7

Industry responses 15

Implications for online safety 24

Conclusions and next steps 28

Annex

A1. Stakeholder Engagement 29

Executive Summary

On 14 May 2022, an attack was carried out in Buffalo, New York, which resulted in the death of ten individuals and the wounding of three others. The attack was livestreamed online and versions of the footage were disseminated on multiple online services, potentially exposing UK users to content related to terrorism. As the regulator of UK established video-sharing platforms (VSPs) and the prospective UK Online Safety regulator, we sought to learn from the tragic event by reviewing industry responses to the livestreamed attack and cross-industry collaboration to prevent dissemination of associated content.

The Buffalo attack showed that terrorist, violent and hateful content online can play a significant role in radicalising vulnerable users. The attacker was reportedly inspired by, and used a similar modus operandi to, the Christchurch mosque attacks in New Zealand, where 51 individuals lost their lives. Large sections of the Christchurch attacker's manifesto were copied and numerous references were made to him in the Buffalo attacker's diary.

Such an attack highlights the complexity and challenges associated with terrorist content online. It has been reported that the attacker was radicalised in part through exposure to racist content on the message board 4chan. He also appears to have acted alone and did not belong to any terrorist or other related organisation. Links to his online diary were shared in private servers and through direct messages approximately 30 minutes before the attack; 15 individuals clicked on the link to the diary. The livestream of the attack lasted less than 2 minutes and there were 28 viewers, or fewer, who watched the livestream of the channel at some point during broadcast. Despite this low figure, footage of the attack was spread across platforms and seen by millions of people, and copies of the diary and manifesto were shared mainly through smaller platforms. The potential of harm from such an exploitation of online services is multifaceted: the disturbing and graphic nature of the footage, the (re)traumatisation of communities who have been affected by similar incidents and the risk of radicalisation of vulnerable online users.

After becoming aware that the attack was livestreamed and being disseminated on video-sharing platforms (VSPs) that we currently regulate, we urgently arranged to meet representatives from the relevant platforms to establish precisely what had happened and how effectively their systems and processes had responded to the livestreamed attack.

Twitch engaged with us proactively to brief us in detail and TikTok and BitChute also cooperated comprehensively with our enquiries. On the basis of thorough assessments of these platforms, and our consideration of whether there was evidence of potential failures to comply, we have decided not to open formal enforcement investigations at this stage. However, we will continue to engage with these three VSPs about the safety measures they have in place to prevent the livestreaming of such attacks and protect users from the dissemination of associated content.

The footage and associated content were also distributed on other services likely to fall within scope of our forthcoming regulation of online safety. To help us develop a comprehensive understanding of the attack, we held over 20 meetings with expert stakeholders including industry actors, researchers, governments and regulators with expertise and experience of online radicalisation and the dissemination of terrorist content online.

We have identified four key challenges and choices facing companies seeking to prevent terrorist exploitation of their services:

- Cross-industry initiatives like the Christchurch Call (CCU), Global Internet Forum to Counter Terrorism (GIFCT), EU Internet Forum (EUIF) and Tech against Terrorism's (TaT) Terrorist Content Analytics Platform have had a notable impact in reducing the dissemination of footage related to livestreamed terrorist attacks since Christchurch, especially through their crisis response protocols. However, no one cross-industry initiative is a panacea to an eco-system wide challenge, especially if membership or participation is limited. Without the inclusion of smaller and newer online services, it may be that no single model can provide a comprehensive solution. Multiple approaches deployed simultaneously, addressing different services and different types of threat, may be increasingly needed.

- Platforms take different approaches to identifying terrorist, violent extremist or hateful actors. Some only prohibit organisations that have been proscribed by states or intergovernmental organisations. Others take a more expansive approach, banning a wider range of groups and content, including hate speech, incitement to violence, and disinformation/conspiracy theories that seek to radicalise users. Services can set their own terms and conditions as long as they meet any specific conditions required by relevant laws and regulation. But in doing so, they inevitably make trade-offs between protecting their users from potentially harmful content, and their ability to say what they want. For users to know what they may or may not share or encounter on platforms, they need clarity in the platforms' terms of service on how platforms define and tackle terrorist, violent and hateful content, and for the terms of service to be accessible to them.

- User reporting tools continue to play an important role in flagging terrorist content and violent and hateful content in livestreaming, and users are most likely to use them when doing so is easy and accessible. Backing this up by appropriately resourced content moderation teams could also strengthen its effectiveness.

- Though livestreamed terrorist attacks are rare, they are uniquely disturbing and pose significant risks to users. There are opportunities for platforms to reduce the risk of these incidents, by introducing features that limit access to livestreaming in particular circumstances. We encourage platforms to examine the risks posed by their services and make appropriate efforts to make themselves robust against exploitation by attackers.

This work has helped build our understanding of the ways in which terrorist and violent extremist actors use livestreaming to disseminate footage of their attacks, and the impact of industry responses and collaborative action. We will take these insights into account as appropriate in our continuing work towards the launch of the UK's Online Safety regulation, including preparation of our sector risk assessment, risk assessment guidance for regulated services and codes of practice relating to illegal content, which we plan to publish in draft form in Spring 2023 for consultation, following passage of the Online Safety Bill. Between now and then, we will continue to work with UK-established VSPs to ensure they have appropriate measures to protect users from illegal terrorist material and content that incites hatred and violence. We also expect to publish further research into risks posed by terrorist, violent and hateful content to UK online users in the coming year.

Introduction

About the attack

On 14 May 2022, an 18-year-old far-right extremist allegedly undertook a violent terrorist attack on a supermarket in a predominantly Black neighbourhood in Buffalo, New York. During the attack, he killed ten individuals and injured three others, the majority of whom were Black. The attack was livestreamed[1], recorded and disseminated on several online services along with a manifesto and 'diary'. Based on subsequent hate crime charges brought against the alleged attacker by the US Justice Department[2], the attack appears to have been racially motivated and drew significant inspiration from previous far-right attackers, including one carried out in Christchurch, New Zealand, in 2019. It was perpetrated by an individual seeking to maximise the spread of footage of their livestreamed attack.

Although the event took place in the US, the content was distributed globally and, to the extent that it could incite violence or hatred and promote terrorism, posed a risk to UK users who viewed it and the wider public. Many of the online services on which the content was hosted or distributed have measures in place to limit the viewership and prevent the spread of terrorist content, and some participate in cross-industry initiatives to support this. However, it is likely other individuals will attempt to livestream similar attacks in the future and continue to spread terrorist content.

The Buffalo attack was not simply an incident with implications for online safety, but one which had a devastating and traumatic impact on the victims, their families, members of the Black community in Buffalo, and the wider Buffalo community. As such, our thoughts are with the victims and their families. Out of respect for those affected, we have decided not to refer to the attacker by name.

Ofcom's role

We are the regulator for video-sharing platforms (VSPs) established in the UK under legislation stemming from the EU Audiovisual Media Services Directive 2018.[3] Under UK law, VSPs must protect all users from material likely to incite violence or hatred against particular groups and content which would be considered a criminal offence under laws relating to terrorism; child sexual abuse material; and racism and xenophobia. These requirements came into effect on 1 November 2020, and since then we have been developing and implementing the regulatory framework.

The Buffalo attack was livestreamed on Twitch, a UK-established VSP with which we have been working since the VSP regulations came into force. Copies of video footage from the livestream were

[1] For a detailed explanation of a livestream, and attacks which have used this functionality in the past, please refer to pages 9-12 of this document.
[2] The United States Department of Justice, July 2022. Federal Grand Jury Indicts Accused Tops Shooter on Federal Hate Crimes and Firearms Charges in Buffalo, New York.
[3] Part 4B, Communications Act 2003.

then shared on other UK-regulated video-sharing platforms, including BitChute and TikTok[4]. Content also spread across other platforms, such as Facebook, Instagram, YouTube, and Streamable.

In the days following the attack, Ofcom engaged with notified VSPs that were impacted to understand the measures they had in place to protect their users from terrorist content. In addition, we met with experts in monitoring, analysing and combating the spread of terrorist content online and other third-party organisations and services – both currently regulated under the VSP regime and those who will likely be in-scope of the Online Safety Bill.

In addition, we engaged with relevant agencies and regulators in the UK and in other jurisdictions with whom we have established relationships. This has involved meetings with the UK Home Office, the Christchurch Call Unit within the New Zealand Government, Australia's eSafety Commissioner and the Global Internet Forum to Counter Terrorism (GIFCT).

The purpose of this report is to describe how footage of this attack, and related material, came to be disseminated online, and to understand the implications of this for efforts to keep people safe online. We will use this insight to inform our future policy under the UK's proposed Online Safety Bill, and our ongoing supervisory engagement with UK-established VSPs. While we have engaged with regulated VSPs to establish how effectively their systems and processes responded to the livestreamed attack, we have not conducted any formal investigation using our existing statutory powers under VSP legislation in relation to any of the matters contained in this report. We have therefore made no findings of fact pursuant to our formal powers. Where we discuss factual matters, we reference whether the information is taken from a public source or direct from the relevant stakeholder.

The UK's Online Safety Bill

The Online Safety Bill[5] will require online services which host user-generated content or are search engines to have systems and processes for protecting individuals from certain types of harm online. Any such service which has significant numbers of UK users or which is targeted at the UK market will have new duties and must comply with the law. The current version of the Bill lists a number of UK terrorism offences as a type of priority illegal harm. Consequently, a wide range of services will, once the Bill passes, have a legal duty to assess the risk of their service being used for the commission or facilitation of these terrorism offences and the risk of users encountering content amounting to such offences ('terrorism content'). Services will also have to take proportionate measures relating to the design and operation of the service to prevent individuals from encountering terrorism content and effectively mitigate and manage the risks they have identified of the service being used for the commission or facilitation of terrorism offences and of harm to individuals. Finally, services will also need to use proportionate systems and processes designed to

[4] TikTok has shared with us that of the livestream content viewed on TikTok in the UK, the majority of that content was non-graphic in nature.

[5] The Online Safety Bill was introduced into Parliament in March 2022. A further version of the Bill, as amended at Committee stage in the House of Commons, was published on 28 June 2022. Further amendments to the illegal content duties in the Bill were made on the first day of report stage on 12 July 2022. This summary is based on the version of the Bill, as last amended on 12 July 2022.

take terrorism content down swiftly when they become aware of it and minimise the length of time for which it is present.

Ofcom will be required to produce a register of risk, risk profiles, guidance for risk assessments and codes of practice on illegal content and offences, which will assist services in understanding how they can comply with their duties. In parallel, we will engage with high-risk or high-impact services to understand their existing safety systems and how they plan to improve them. The current Bill would give us extensive information gathering powers and we may use these if needed to gather evidence for our work on implementing the regime.

As appropriate, we will take account of the insights and issues arising from this project in our future policy work, including our codes of practice relating to illegal content, which we currently plan to publish in draft form for consultation in Spring 2023, following passage of the Online Safety Bill.

Background to the attack

Radicalisation and the role of online platforms

Although it can be challenging to build a complete picture of an individual's radicalisation journey, so-called 'digital footprints' allow expert researchers, organisations and law enforcement agencies to create a picture of a user's radicalisation pathway. The alleged Buffalo attacker appears to have left behind a vast cache of online content that has since been analysed and reportedly provides insights into what may have led to the attack.[6]

Research conducted since the attack points to similarities between other far-right terrorist attacks, such as the publication of a manifesto.[7] The 180-page manifesto was published on Google Drive and reportedly sets out the attacker's ideology and reasons for undertaking the attack.[8]

The attacker reportedly maintained a private server on the platform Discord, which documented many of his thoughts and intentions.[9] It contained 673 pages of logs and has since been referred to as the 'diary'.[10] The attacker invited individuals to the private server approximately 30 minutes before the attack took place; only 15 individuals clicked through to the private server.[11]

Both the manifesto and diary contained several explicit references to far-right themes and conspiracy theories, as well as racist and antisemitic content. This included 'The Great Replacement' conspiracy theory, which claims White western populations are being systematically replaced by 'non-whites'.[12] Many antisemitic, far-right proponents of the theory believe that this is being deliberately orchestrated by Jewish people.

Within the manifesto and diary, the attacker is reported to have stated it was "boredom" during the early period of the COVID-19 pandemic in 2020 that led him to the imageboard site, 4chan[13], and specifically to boards devoted to guns.[14] Analysis of the manifesto and diary suggested that spending time on 4chan led the attacker to more extreme spaces, focused on social issues and politics where he was introduced to violent extremist, racist and antisemitic content, specifically the so-called

---

[6] Combating Terrorism Center (Amarnath Amarasingam, Marc-André Argentino, Graham Macklin), July 2022. The Buffalo Attack: The Cumulative Momentum of Far-Right Terror.

[7] International Centre for Counter-Terrorism (ICCT) (Prof. Tahir Abbas, Inés Bolaños Somoano, Joana Cook, Isabelle Frens, Graig R. Klein, Richard McNeil-Willson), May 2022. The Buffalo Attack – An analysis of the Manifesto.

[8] CTC, The Buffalo Attack: The Cumulative Momentum of Far-Right Terror.

[9] ICCT, The Buffalo Attack – An analysis of the Manifesto.

[10] Global Network on Extremism and Technology (Laurence Bindner, Raphael Gluck), July 2022. The Buffalo Attack – Insights from the Suspected Terrorist's Diary. The Wall Street Journal, May 2022. Buffalo Shooter's 673-Page Diary Reveals Descent Into Racist Extremism.

[11] Discord, May 2022. Our response to the tragedy in Buffalo.

[12] ICCT, The Buffalo Attack – An analysis of the Manifesto.

[13] 4chan is an online 'imageboard' forum in which users can post comments and share images and videos. The forum was modelled on Japanese imageboards to discuss anime (Japanese animation) and share pornographic imagery and other types of fringe content. The site is responsible in-part for popularising internet memes and has attracted several controversies related to content and behaviour on the site. This includes the gamergate scandal (Wired, September 2014. How 4chan manufactured the #GamerGate controversy), as well as allegedly hosting racist and antisemitic content – The Guardian, May 2022. How 4chan's toxic culture helped radicalize Buffalo shooting suspect.

[14] GNET, Insights from the Suspected Terrorist's Diary.

'politically incorrect' message board.[15] In the diary, the alleged attacker claims he became convinced of the claims he encountered on 4chan and that it was on this site that he first encountered footage from the Christchurch attacks.[16]

Many expert researchers, including some those Ofcom spoke to for the purpose of this report, point to the radicalising effect of previous high-profile attacks on those who have yet to carry out violence.[17] Likewise, they highlight the crucial role of certain online spaces, communities, and hateful material in an individual's radicalisation process. It is also important to note that radicalisation is not purely a consequence of the online sphere but often relies upon multiple online and offline factors.[18]

In this case, the alleged attacker appears to have been heavily inspired by previous extreme far-right terrorists, specifically the Christchurch 2019 attacker. Reports suggest the Buffalo attacker followed a very similar *modus operandi* to the Christchurch attacker, posting a manifesto online shortly before the attack and livestreaming throughout using a camera attached to a tactical helmet.[19] In fact, large parts of the Buffalo attacker's manifesto were heavily plagiarised from the Christchurch attacker's manifesto.[20]

According to the manifesto, the alleged attacker began thinking of carrying out an attack in November 2021, having spent over a year on 4chan forums. In that month he wrote on a 4chan message board that 'a [refers to the name of the Christchurch attacker] event will happen again soon'.[21]

The diary also outlines a set of reasons for utilising Twitch for the livestream. The alleged attacker indicated an initial desire to use Facebook, like the Christchurch attacker, but decided to use Twitch 'because only boomers[22] actually have a Facebook account'.[23] The diary also noted that Facebook requires viewers to have an account, as opposed to Twitch which only requires users to create an account to livestream.[24] This contrasts with other services, such as YouTube, which require a user to have a minimum number of subscribers to stream from a mobile device.[25]

This highlights the careful thought the author of the diary put into considering the platform best suited to his intentions. The diary documented extensively how to use the livestream, including potential issues that may arise. These included bandwidth issues and how to resolve them, and

---

[15] GNET, Insights from the Suspected Terrorist's Diary.
[16] GNET, Insights from the Suspected Terrorist's Diary.
[17] For example, see US Department of Homeland Security, August 2020. Mass Attacks in Public Spaces, page 20; Institute for Strategic Dialogue (Jacob Davey and Julia Ebner), July 2019. 'The Great Replacement': The violent consequences of mainstreamed extremism.
[18] Ministry of Justice Analytical Series (Dr Jonathan Kenyon, Dr Jens Binder, Dr Christopher Baker-Beall), September 2021. Exploring the role of the Internet in radicalisation and offending of convicted extremists.
[19] CTC, The Buffalo Attack: The Cumulative Momentum of Far-Right Terror.
[20] CTC, The Buffalo Attack: The Cumulative Momentum of Far-Right Terror.
[21] Washington Post, May 2022. Buffalo shooting suspect wrote of plans 5 months ago, messages show; The Guardian, May 2022. Buffalo shooting: gunman plotted attack for months.
[22] A boomer, a reference to baby boomers, is a popularised term to refer to someone older, typically in a derogatory, satirical and/or humorous manner. In this instance it was used to refer to the perceived age range of users of Facebook.
[23] CTC, The Buffalo Attack: The Cumulative Momentum of Far-Right Terror.
[24] The New York Times, May 2022. After Buffalo Shooting Video Spreads, Social Platforms Face Questions.
[25] YouTube's mobile live streaming requirements state: 'To live stream on mobile, you'll need: At least 50 subscribers. If you're between the age of consent to 17, you need at least 1,000 subscribers to live stream from a mobile device.' YouTube Help, 2022. Create a live stream on mobile [accessed 4 October 2022].

connectivity problems between a phone and a GoPro, for example.[26] In January 2022, the attacker reportedly started planning the attack more seriously, buying surplus military gear and ammunition.[27]

What happened in Buffalo?

The attacker livestreamed his actions through his channel on Twitch. The livestream lasted under 25 minutes[28], with the violence taking place only at the end. Twitch terminated the livestream within two minutes of when the shooting began.[29] Despite its removal, it is reported the footage of the shooting was captured and subsequently disseminated across a range of platforms including on Twitter[30], Facebook[31], Streamable, TikTok[32], BitChute, WhatsApp, Instagram and Telegram.[33]

It is likely, although unproven, that a viewer of the livestream captured a recording of the footage and then distributed the footage, enabling its proliferation across multiple platforms. It is unclear who captured the footage and what process they used to do so.

What is livestreaming?

Livestreaming is online streaming media that is simultaneously created and broadcast in real time. Although the content of livestreams can differ vastly, ranging from adult content to gaming, the core characteristics of a stream will remain largely the same. Typically, particularly in relation to user-to-user livestreaming services such as social media, there will be ways in which viewers and streamers can interact alongside the stream, such as via a live chat function.

Livestreaming and terrorism

Whilst the use of video by terrorist groups and individuals is not a new phenomenon, the use of livestreaming functionalities provided by online platforms is a recent innovation. This has largely come about due to the now widespread accessibility of livestreaming functionalities online. Most recently, livestreaming has been used by far-right terrorists to broadcast attacks live from a first-person perspective. In most cases, live viewership has been relatively low, however the subsequent recording and dissemination of the footage has led to its virality.[34]

---

[26] CTC, The Buffalo Attack: The Cumulative Momentum of Far-Right Terror.
[27] The Buffalo news: May 2022. Timeline: What happened prior to Saturday's mass shooting in Buffalo.
[28] The Washington Post, May 2022. Only 22 saw the Buffalo shooting live. Millions have seen it since.
[29] Twitch, 2022. A statement from Twitch regarding the Buffalo supermarket hate crime.
[30] Twitter suggested to us that the capturing and subsequent dissemination of the footage was carried out by sympathisers of the attacker.
[31] Facebook provided further nuance in that it was not the footage itself of the attack that was broadly disseminated on Facebook or Instagram. Rather, what was primarily disseminated were URL addresses to third-party hosted content.
[32] TikTok clarified to us that that the majority of the content associated to the livestreamed attack viewed on TikTok in the UK was non-graphic in nature.
[33] The Washington Post, May 2022. Only 22 saw the Buffalo shooting live. Millions have seen it since.; Metro, May 2022. Buffalo shooter's livestream video broadcast on social media millions of times; Anti-Defamation League, May 2022. Footage of Buffalo Attack Spread Quickly Across Platforms, Has Been Online for Days.
[34] Combating Terrorism Center (Graham Macklin), July 2019. The Christchurch Attacks: Livestream Terror in the Viral Video Age; New York Times, October 2019. 2,200 Viewed Germany Shooting Before Twitch Removed Post.

Research has found that livestreamed terrorist attacks have played a particularly prominent role in radicalising other users.[35]

Christchurch attacks

The attack in Buffalo and the subsequent livestream was directly inspired by an attack that took place in Christchurch, New Zealand in 2019, where 51 individuals were killed in an anti-Muslim terrorist attack livestreamed on Facebook.[36] The video had garnered approximately 4,000 views before removal. Similar to the Buffalo attack, the Christchurch attack had been signposted on a fringe unmoderated chat board – in this case, 8chan. This same service was then allegedly used to share a link to the footage; in the 24 hours that followed, Facebook removed 1.5 million videos of the attack and a further 1.2 million were blocked at upload.[37]

Halle and other attacks

This is not the only time that a livestreaming feature has been abused (or attempted to be abused). In 2016, a man in France used the Facebook Live feature to broadcast his justification for killing two police officers whilst holding a 3-year-old hostage and pledging his allegiance to the Islamic State.[38] In 2019, a man attempted to use Facebook Live whilst attacking a Synagogue in Poway, California, USA, killing one woman, but reports suggested there was a malfunction.[39] In 2019 a gunman reportedly livestreamed himself through his channel on Twitch attacking a synagogue and a kebab shop in Halle, Germany.[40] In 2020, an attacker livestreamed himself on Snapchat carrying out an attack in a mall in Glendale, Arizona.[41] Following these attacks, many online services put stronger measures in place to identify and remove terrorist content from their platforms and limit the number of users who will see the content. For example, following the Christchurch attack, Facebook implemented several restrictions on its livestreaming functionality. This included a 'one strike policy' that strengthened the rules on using the Facebook Live feature.[42]

[35] ISD (Jacob Davey and Julia Ebner), 2019. 'The Great Replacement': The violent consequences of mainstreamed extremism

[36] The Guardian, May 2022. Buffalo shooting: unease in New Zealand as live stream of 'Christchurch-inspired' attack finds foothold.

[37] Combating Terrorism Center (Graham Macklin), July 2019. The Christchurch Attacks: Livestream Terror in the Viral Video Age; Meta Newsroom, March 2019. Update on New Zealand.

[38] The Verge, June 2016. French terror suspect reportedly streamed attack on Facebook Live.

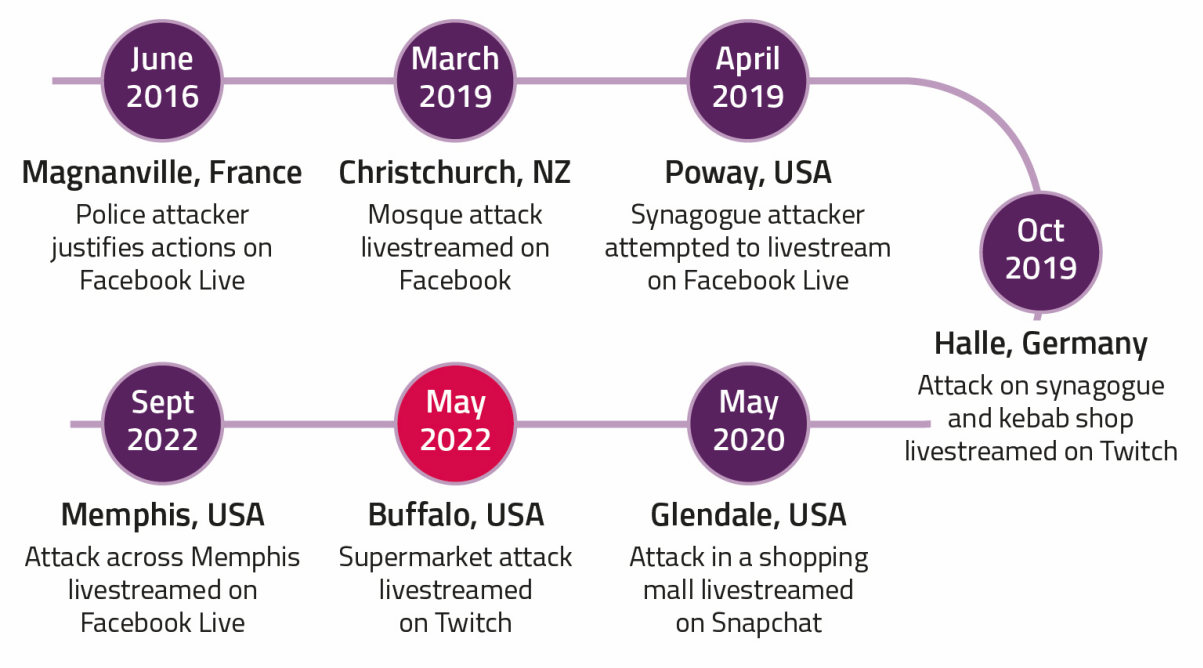
[39] ABC News, April 2019. Alleged San Diego synagogue shooter John Earnest had 50 rounds on him when arrested: Prosecutor.

[40] BBC News, December 2020. Halle synagogue attack: Germany far-right gunman jailed for life.

[41] Vice News, May 2020. A mass shooter live-streamed his attack on snapchat at an Arizona mall.

[42] Meta, May 2019. Protecting Facebook Live From Abuse and Investing in Manipulated Media Research. Meta also shared the following additional point on the changes made following the Christchurch mosque attacks: i) restricting users if they have violated certain rules e.g. the Dangerous Organizations and Individuals policy, and ii) updating proactive detection systems and reducing the average time it takes for AI to find a violation on Facebook Live.

Figure 1: Recent incidents with a livestream element



Source: See references 38 to 42.

Memphis

In September 2022, an individual carried out seven shootings in Memphis, Tennessee, United States, leaving four people dead and three injured.[43] The alleged perpetrator livestreamed their conversations and part of the attack using Facebook Live.[44] It has been commented that this does not appear to have been an ideologically motivated incident.[45]

As with the Buffalo attack, GIFCT activated its Content Incident Protocol[46] in response to the attack, as it was deemed that copies of the livestream, or parts of the livestream, which showed the violence spread across several online services. GIFCT enabled members to be able to share hashes[47] of the perpetrator-produced content to its hash-sharing database for members to address in accordance with their respective policies.[48]

Analysis carried out by one of the organisations we met suggests that footage of the livestream has not spread to the same extent as that from Buffalo. However, insights

[43] The Daily Beast, September 2022. Live-Streaming Gunman in Custody After Memphis Locked Down.
[44] Washington Post, September 2022. 4 killed, man arrested in Memphis shootings after grisly live stream.
[45] Global Network on Extremism and Technology (Sammie Wicks), October 2022. Nihilism and Mass Shooterism: unclear categories and potential dangers
[46] For a detailed explanation of the GIFCT and its Content Incident Protocol, please see pages 16-17 and 19.
[47] For a detailed explanation of hashes, please see page 20.
[48] GIFCT, September 2022. Content Incident Protocol Activated in Response to Shooting in Memphis, Tennessee, United States.

shared with us from Memetica[49] suggest the footage has been shared within far-right networks.[50] Tech Against Terrorism stated that 'this incident has inflamed long-standing narratives among some violent far-right extremists of an ongoing "war on white people"; we've seen several instances of Europe / North America-based far-right extremists describing it as an "anti-white" attack, citing no evidence.'[51]

When compared with the greater spread of content from the attack in Buffalo, this suggests that ideology and promotion by a small but dedicated group of followers can make a significant difference to the dissemination of content. Seemingly, very little effort is needed to generate global coverage of an event; only 28 individuals, or fewer, watched the livestream of the channel at some point during broadcast of the livestreamed attack in Buffalo and copies of this had significant spread. Industry responses appear to have played very little role in determining spread: GIFCT activated its Content Incident Protocol in both cases, yet Buffalo content achieved much more widespread distribution than Memphis. For all these reasons, the potential to cause harm can vary significantly for similar types of content.

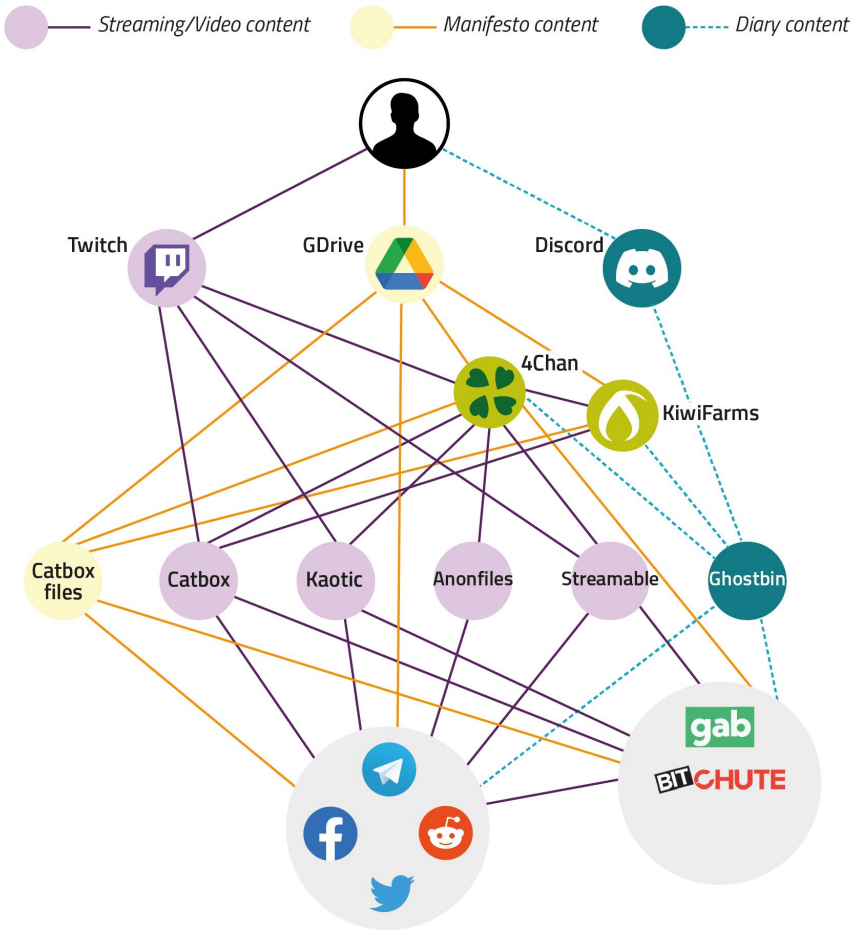
Subsequent dissemination of content online

Like previous extreme far-right terrorist attacks, it appears one key aim of the alleged Buffalo attacker was to ensure that his own content, including the livestream, manifesto and diary were disseminated widely – ensuring his own legacy, contributing to the online extreme far-right ecosystem and potentially radicalising others towards violent action. Figure 2 below demonstrates how this content spread across numerous platforms, including both larger and smaller platforms. Evidence also suggests that links to this content continued to be shared in the days following the attack.[52]

The attacker did not plan for how content would get disseminated. In fact, the evidence suggests that only a small number of users initially engaged with the content.[53] Rather, it is far more likely that the dissemination of content relied on those users' existing networks, as well as the attacker's assumption that controversial, highly shocking and 'relevant' content would organically spread online, as it had done for previous similar attacks.

[49] Memetica is a digital investigations group providing intelligence and risk advisory services on a variety of strategic issues relating to coordinated harassment, violent extremism, and disinformation.
[50] This can also be seen in GNET, Nihilism and Mass Shooterism: Unclear Categories and Potential Dangers.
[51] Tech Against Terrorism analysis of content associated with the attack in Memphis.
[52] Tech Policy Press (Adi Cohen, Benjamin T. Decker), May 2022. Internet Trolls Should Not Dictate the Terms of Public Exposure to Hate.
[53] Discord stated that 15 users clicked an invitation sent in the 30 minutes leading up to the attack to join the perpetrator's private Discord server where they may have viewed his personal diary.

Figure 2: Memetica analysis of the dissemination of content related to the Buffalo attack.[54]



Source: Tech Policy Press (Adi Cohen, Benjamin T. Decker), May 2022, Trolls should not dictate the terms of public exposure to hate.

The spread of such content has real-life impact offline, including potentially inspiring further far-right attacks. One of the research organisations we interviewed, Moonshot[55], found in its research a 186% increase in discussion of conspiracy theories, such as the 'Great Replacement Theory' and 'White Genocide', within online US violent extremist spaces following the attack.[56] Further, in August 2022, the U.S. Department of Homeland Security (DHS), Federal Bureau of Investigation (FBI) and the National Counterterrorism Center (NCTC) assessed that the manifesto will 'likely enhance the capabilities of potential mass casualty shooters who may be inspired by this attack'.[57]

[54] With Memetica's approval, we have used their analysis for the graph. We have not repeated Memetica's analysis of these services, nor engaged with all of those mentioned in this graph.

[55] Moonshot is a social impact company which builds solutions to understand and prevent online harm including violent extremism.

[56] Moonshot, May 2022. Moonshot Threat Bulletin: National Trends (May).

[57] CBS News, August 2022. FBI, DHS issue bulletin warning of potential for racially motivated copycat attacks.

The dissemination of terrorist content as well as violent and hateful content across the internet presents a twin challenge to online platforms, both large and small. The first challenge is becoming aware of the content and removing it quickly. The second challenge is the ability for platforms to keep up with the rate at which content is shared – in other words, the need to remove content at a faster rate than it spreads. This second challenge is made more complicated by actors seeking to maximise its distribution through tactics in order to increase its virality or to evade platform detection. Such tactics can, in some instances, provide challenges for automated content detection and enforcement tools that are deployed by platforms, such as hashing and hash matching.[58]

[58] See information box on page 20.

Industry responses

Most platforms have systems, processes and threat detection measures in place to assure removal of content breaking their own terms and conditions. In addition, there are multiple initiatives which aim to improve cross-industry responses to terrorist content and activity online. Below are some examples of these.

Cross-industry Initiatives

The Christchurch Call

Following the terrorist attack in Christchurch, New Zealand and France launched the Christchurch Call. Governments and online service providers that support the Call commit to various actions to prevent and eliminate terrorist and violent extremist content online, working collaboratively with civil society and upholding human rights and a free, open and secure internet.[59]

The Christchurch Call contains a number of commitments to prevent and eliminate terrorist and violent extremist content online. There is a specific commitment for governments and online service providers to work together to enable a rapid and coordinated response to an online incident like the livestreamed attacks in Christchurch.[60] The Christchurch Call Crisis Response Protocol was the first overarching initiative to establish what constitutes a crisis, the actions different sectors should take and how to communicate within the Call Community during a crisis. Other protocols, including the GIFCT Content Incident Protocol described later in the document, were developed to work in conjunction with the Crisis Response Protocol.

The European Union Internet Forum (EUIF)

The EUIF was launched by the European Commission in December 2015. Its mission is to 'provide a collaborative environment for governments in the EU, the internet industry, and other partners to discuss and address the challenges posed by the presence of malicious and illegal content online'.[61] One of its key areas of work was being involved in the creation of Europol's EU Internet Referral Unit (EU IRU); its aim is to 'combat terrorist propaganda and related violent extremist activities on the internet'.[62] The EU IRU identifies and refers relevant content to the service providers in order to review and remove content which breaches their terms and conditions.

[59] Christchurch Call. Christchurch Call Story.
[60] Christchurch Call. Crisis and Incident Response.
[61] European Commission, European Union Internet Forum (EUIF) [accessed 5 October 2022].
[62] Europol, 2022. Europol's Internet Referral Unit to combat terrorist and violent extremist propaganda.

The EUIF currently has 23 members[63] encompassing EU countries[64], European institutions and agencies[65], online services[66], GIFCT, Tech against Terrorism, United Nations entities[67], Radicalisation Awareness Network, the WeProtect Global Alliance and Tech Coalition. It also has agreed its own EU Crisis Protocol 'to respond to the viral spread of terrorist and violent extremist content online'[68], which can be activated by any member state or Europol. To date, it has only activated once; when content associated with the terrorist murder of Samuel Paty was circulated.[69]

GIFCT

GIFCT is another cross-industry initiative designed to prevent terrorists and violent extremists from exploiting digital platforms.[70] The initiative was founded in 2017 by Facebook, Microsoft, Twitter and YouTube, becoming a non-governmental organisation (NGO) in 2019, and states its mission as preventing terrorists and violent extremists from exploiting digital platforms.[71]

Following criticism of platforms after the Christchurch attack, GIFCT's efforts evolved. It became an independent NGO with its own dedicated technology, counterterrorism and operations teams, committing to end terrorist abuse of member platforms. It has also introduced greater multi-stakeholder involvement through its Working Groups and Independent Advisory. GIFCT has since grown its membership, developed new technologies which are shared amongst members and offers mentorship programmes to those wishing to join, to improve safety on platforms. Its most recent and 19th member was Clubhouse.[72]

It has four goals[73] :

- Empower a broad range of technology companies, independently and collectively, with processes and tools to prevent and respond to abuse of their platforms by terrorists and violent extremists;
- Enable multi-stakeholder engagement around terrorist and violent extremist misuse of the Internet and encourage stakeholders to meet key commitments consistent with the GIFCT mission;
- Promote civil dialogue online and empower efforts to direct positive alternatives to the messages of terrorists and violent extremists;
- Advance broad understanding of terrorist and violent extremist operations and their evolution, including the intersection of online and offline activities.

---

[63] European Union Internet Forum (EUIF) [accessed 05 October 2022].
[64] This includes EU countries and countries of the European Free Trade Agreement.
[65] Such as Europol, Eurojust, fundamental Rights Agency, European External Action Service, Council's Counter Terrorism Coordinator.
[66] This includes Automattic, Discord, Dropbox, Meta, Google, Internet Archive, Just Paste.it, Mega, Microsoft, Snap, Telegram, Twitter, Twitch, TikTok, Roblox, Zoom.
[67] United Nations Office of Counter-Terrorism and United Nations Security Council Counter-Terrorism Committee.
[68] European Commission, October 2019. Fighting Terrorism Online: EU Internet Forum committed to an EU-wide Crisis Protocol.
[69] Speech by President von der Leyen on the occasion of the Christchurch Call Second Anniversary Summit.
[70] GIFCT. About.
[71] GIFCT. About.
[72] GIFCT, September 2022. Expanding our Collective Capacity: GIFCT's Progress Continues.
[73] GIFCT. About.

Some experts interviewed for this report felt that GIFCT's shared hashing database and CIP have significantly contributed to the reduction in volume and spread of terrorist images and videos across its member platforms since the Christchurch attack. Some experts interviewed also suggested that GIFCT could improve in certain areas, some of which it is working to address. While its membership continues to expand, its hashing database is not available to every industry actor, only its members. GIFCT told us that potential members need to demonstrate a level of trust and shared purpose framed as a commitment to a culture of multi-stakeholder collaboration and respect for human rights.

This is reflected in GIFCT's eligibility criteria for new applicants, which requires demonstration of i) terms and conditions that explicitly prohibit terrorist and/or violent extremist activity, ii) the ability to receive, review, and act on both reports of activity that is illegal and/or violates terms and conditions and user appeals, iii) a desire to explore new technical solutions to counter terrorist and violent extremist activity online, iv) regular, public data transparency report, v) a public commitment to respect human rights in accordance with the United Nations Guiding Principles on Business and Human Rights, and vi) support for expanding the capacity of civil society organisations to challenge terrorism and violent extremism. GIFCT emphasised to us that beyond the above criteria and a willingness to join, there are no barriers to membership. For those companies that are willing to join but do not satisfy all the criteria, they may become eligible by completing TaT's mentorship programme.

A further challenge, faced by all shared hash databases, is the risk of false positives: the inclusion of material wrongly deemed to be terrorist in nature. While any content classification system will contain errors, the impact of these is magnified if the same database is used to enforce policies on terrorist content across multiple sites. Organisations committed to accountability of such systems are likely to take robust steps to minimise errors, regularly audit effectiveness, and be transparent about error rates and action to reduce these. As such, it is reassuring to see much of this reflected in the recommendations of the Human Rights Impact Assessment of GIFCT conducted by BSR[74] and the GIFCT publishing its own transparency reports. On governance and accountability, BSR particularly notes that 'GIFCT's Operating Board, which currently consists of four founding member companies, is not a sustainable model over the medium-to-long term and recommend that GIFCT consider the merits of transitioning to a multi-stakeholder decision-making model in two years.'

Tech against Terrorism

Tech against Terrorism (TaT) is an initiative launched and supported by the United Nations Counter Terrorism Executive Directorate (UN CTED),[75] committed to 'supporting the tech industry [to] tackle terrorist exploitation of the internet, whilst respecting human rights'.[76] In November 2020, TaT launched the Terrorist Content Analytics Platform (TCAP), which TaT describes as the world's largest database of verified terrorist content. Through this, TaT is able to notify platforms of the presence of terrorist content on their platforms.[77]

[74] A Human Rights Assessment of the GIFCT. BSR Blog.
[75] Tech Against Terrorism. About.
[76] Tech Against Terrorism. Home.
[77] Tech Against Terrorism, 2021. Terrorist content analytics platform.

The TCAP is a database of URLs or websites containing terrorist content. If TaT finds a URL carrying illegal terrorist content on a particular platform, it is sent an alert. The concerned platform then decides whether the content violates its policies. This service is provided both to smaller services which may not have the resources or technology to track URLs down themselves, and to larger platforms. TaT's coverage is not limited to its members – it is able to contact any platform to notify it of URLs in the TCAP.

TaT flags URLs containing terrorist content to more than 100 platforms. Its platform includes content from groups, individuals or attacks that have been designated as terrorist by democratic states and intergovernmental organisations. According to TaT, this approach helps ensure that they only flag content that has been produced by appropriately designated entities. In crisis situations, TAT overrides the TCAP's inclusion policy to include crisis material such as the livestream and manifesto produced by the perpetrator of the attack in Buffalo. This process applies during crisis situations initiated by the three major crisis protocols.

Response from industry

As demonstrated throughout the previous chapter, a wide variety of online services were used to disseminate content associated with the attack across the internet.

Some platforms where the content was distributed are currently regulated by Ofcom, and they provided greater detail on their response, which is reflected below. Most services, however, will only become regulated by Ofcom when the Online Safety Bill comes into effect. Due to this, we relied on these services' willingness to engage with us and speak to us openly about their response. The information provided within this chapter is therefore based on a number of sources, including the platforms themselves, interviews with a number of expert stakeholders including industry actors, researchers, governments and regulators with expertise and experience of online radicalisation and dissemination of terrorist content online, and cross-industry initiative findings. We believe these interviews and reports provide valuable information about how industry actors responded to the incident, in real-time and in the days that followed.

The livestream

Twitch confirmed to us that the livestream was taken down less than two minutes after the violence began.[78] Shortly before the attack, the attacker sent a link of the livestream to a number of individuals. There were only 28, or fewer, viewers of the livestream, based on the number of video plays triggered during the broadcast. Twitch's moderation practices involve a multi-layered approach including proactive detection, 24/7 review and urgent escalation of user reports and channel moderation measures, which assist Twitch with taking quick action. Soon after the termination of the attacker's livestream, Twitch cooperated with law enforcement agencies and offered assistance. Twitch also proactively reached out to us with a briefing on the incident and its response.

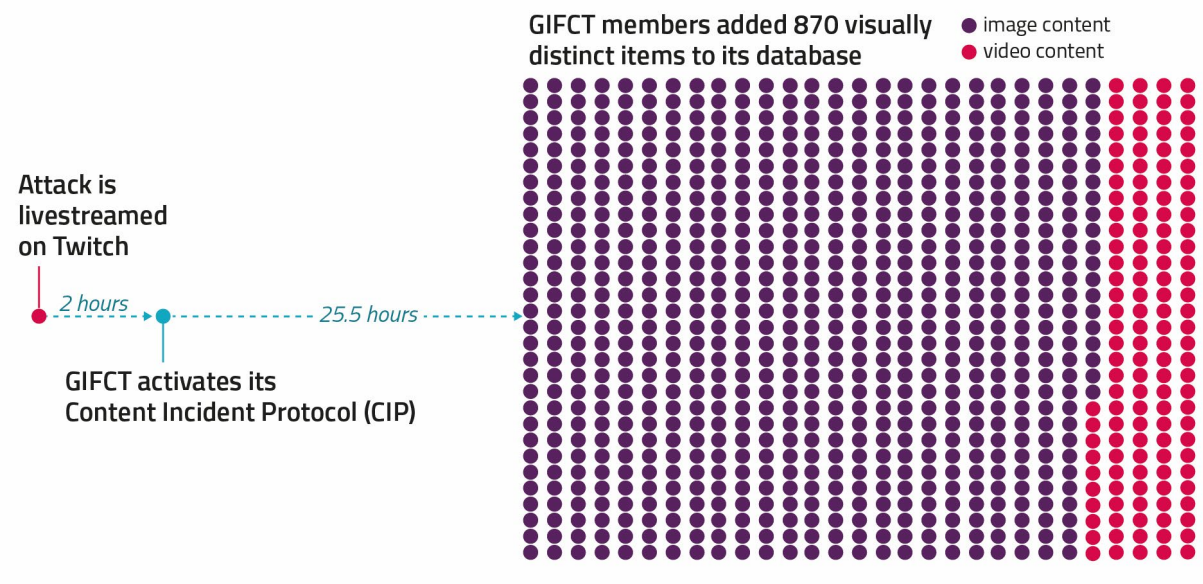
[78] Twitch, 2022. A statement from Twitch regarding the Buffalo supermarket hate crime.

Uploaded versions of the livestream, manifesto and diary

Despite a quick response from Twitch, the livestream was captured and shared across different platforms and services. It is unclear how the footage was captured and how or where it was initially shared. However, Twitch confirmed to us that none of its native on-platform tools were used to download or make copies of the footage.

GIFCT activated its Content Incident Protocol (CIP) approximately 2 hours after the attack in Buffalo took place. Its CIP is activated when a terrorist, violent extremist or mass violence event (1) takes place; (2) is livestreamed; (3) depicts murder or attempted murder; and (4) is being distributed on GIFCT member platforms or so broadly online that such distribution appears inevitable.[79] Once this CIP was activated, and all members were notified, hashes of the attacker's video and related content which could be hashed were shared in the GIFCT hash database. Platforms are able to use these hashes to track down any content on their service and remove it. Between the time the CIP was activated and its conclusion (25.5 hours later), GIFCT members had added approximately 870 visually distinct items (740 images, 130 videos) to its database.[80] As part of its debrief process, GIFCT published a blog post[81] on the activation of the CIP in response to the shooting in Buffalo for further details.

Figure 3: Between when GIFCT activated the CIP and its conclusion, GIFCT members added approximately 870 visually distinct items to the GIFCT hash-sharing database



Source: GIFCT, June 2022. Debrief: CIP Activation, Buffalo, New York USA.

[79] GIFCT. Content Incident Protocol.
[80] GIFCT, May 2022. Update: Content Incident Protocol Activated in Response to Shooting in Buffalo, New York United States.
[81] GIFCT, June 2022. Debrief: CIP Activation, Buffalo, New York USA.

The New Zealand Government's Christchurch Call Unit also convened a debrief on the Buffalo attack with community representatives (supporter governments, online service providers, members of its civil society Advisory Network and partner organisations like the GIFCT and Tech Against Terrorism).

Hashing

Hashing is an umbrella term for techniques to create a short identifier for a file on a computer system. Such files can be images, videos, music, word documents, executables, or any file on a computer system. Hash matching in relation to online content moderation relates to the conversion of mainly video, image and audio content into identifiers or digital 'fingerprints' typically consisting of an alphanumeric string.

These unique identifiers, or hashes, are stored, shared, and then used to identify the same or visually similar content for review and removal. Processes to automatically remove content that matches a previously stored hash are often referred to as hash matching. This can help apply content moderation actions more consistently and at scale but can also amplify the impact of incorrect content moderation decisions. In the case of terrorist content, a video violating a platform's terms of service could be hashed and used to remove copies where it appears.

Hashing is used widely across a number of areas, including copyright, pornographic content and child sexual abuse imagery.[82]

These systems require rapid responses to identify illegal content when it is posted to keep up with the adversarial nature of those seeking to spread this content. The sharing of hashes across online services is also useful in limiting the cross-platform spread of content.

Following the Buffalo attack, URLs were posted on larger platforms which linked back to smaller, less well-known platforms hosting footage of the attack. This allowed duplicates and edits of the livestream to be accessed by users for longer. These platforms tend to not be members of GIFCT, and therefore do not have access to the hash database. TaT, through its TCAP, responded to this challenge and by 19 May had identified 105 unique copies of the livestream and manifesto across 26 online platforms.[83]

TikTok, a platform we regulate under the VSP regime, told us about its response to the Buffalo footage. TikTok prohibits uploads of terrorist content or any content considered to be harmful and has mechanisms in place to report such content. It has an internal crisis incident protocol which was launched following the attack. This included proactive monitoring of relevant content and increased resources on responding to the event. TikTok is not a member of GIFCT but has recently announced

[82] Journal of Online Trust and Safety (Hany Farid), October 2021. An Overview of Perceptual Hashing. See, for example, YouTube, September 2010. YouTube Content ID.
[83] Confirmed by Tech Against Terrorism.

that it is partnering with TaT and is adhering to its membership requirements, including commitments to explore new technical solutions.[84]

Platforms with closed user groups or channels85 or which offer direct messaging between users may allow users to report content that breaches their terms of service, but often do not deploy automated content moderation systems to monitor ~~in~~ what they deem 'private' channels. Some interviewees observed that this may lead to terrorist content remaining on such channels for longer than it might do on public channels. They also believed that footage of the attack, the manifesto and diary are likely to have been shared across these channels, but pointed out it would be difficult to prove this.

While direct messaging services have conventionally enabled communication between two users or small groups, now some, such as Telegram, can facilitate interaction between up to 200,000 users in such closed user groups.[85] This can lead to a large number of individuals seeing illegal content rapidly.[86] However, Telegram told us that in relation to the Buffalo Attack, moderators have been thoroughly removing the video, assessing user reports, and proactively searching for illegal content in public spaces. There is also some evidence, for example from WhatsApp, that automated systems can be deployed on direct messaging services and those that enable closed user groups to detect harmful material on elements such as group photos and names or profile images, alongside network analysis once bad actors are identified.[87]

The diary was initially kept in the attacker's private server on Discord from which he sent invites to view it. Discord released a response to the tragedy online explaining the steps it took after the attack. In our engagement with the platform, it confirmed that the alleged attacker only shared invitations to view his private server approximately 30 minutes before the attack. Fifteen of the recipients clicked on the invitation and would have had access to the diary. Discord acknowledges that subsequent to the attack, there have been reports that the alleged perpetrator was active on public servers, but it states it found no reference to the attack outside of the private server. However, Discord acknowledges that it did not receive any reports about the attacker's activity or the contents of his private server at any point prior to the attack.

Following the attack, Discord banned the attacker's account and removed the diary, working with law enforcement agencies as well as collaborating with the GIFCT. It also says it put measures in place to prevent the spread of content related to the attack, although these are not specified. Discord told us that it is investing in trust and safety and currently has a three-tiered moderation system comprising user controls, Discord's internal moderation systems and processes, and community moderation. It uses third party hashes and is evaluating possible integrations with third party databases, while also investing in internal hashing data. It highlighted that it undertakes

---

[84] TikTok, September 2022. Partnering to prevent violent extremism.
[85] Telegram's FAQs state that groups on its service can have up to 200,000 members each [accessed 3 October 2022].
[86] Telegram have clarified to us that although it is technically possible to create a private community for 200,000 people, as a rule, this does not happen and such groups are mostly public. Unlike many social networks Telegram has no amplifying algorithms for boosting content and users only receive the content they explicitly subscribe for. A private community has no feasible chance of growing beyond a certain number of members far below 200,000. Thus, large channels and groups are public, and for them, have proactive moderation methods in place.
[87] WhatsApp, March 2022. WhatsApp Privacy Policy [accessed 3 October 2022].

proactive moderation efforts on servers, as well as public forums. Such efforts led to the majority of server removals in Q1 2022 for violations of Discord's Terms of Service and Community Guidelines[88].

We also engaged with BitChute, another platform we regulate under the VSP regime, whose terms of service prohibit such content on its platform. Following the attack, BitChute made a statement on its Twitter account asking users to be mindful of its guidelines when reporting on the event. During our engagement, it pointed to benefitting from working with TaT, including being alerted of terrorist content and sharing URLs through TaT's TCAP. We will continue to work with BitChute to ensure it has appropriate measures to protect users from illegal content and incitement to hatred. On 7 October 2022, BitChute announced it was now an official member of TaT.[89]

Response from other stakeholders

Although platforms have a key role in limiting the virality of content online, there are also other stakeholders who play a crucial role following an attack of this nature. Ofcom interviewed research organisations, other regulators and domestic and international government bodies/agencies to gather insights from their perspective on attacks of this nature.

Where a terrorist attack occurs in the UK with an online element (for example, a terrorist attack is livestreamed online), the UK Government will enact its crisis response protocol. This involves working closely with the Counter Terrorism Internet Referral Unit, affected tech companies, and the Global Internet Forum to Counter Terrorism to ensure action is taken against terrorist content related to the attack. The UK Government also works closely with international partners in relation to attacks that take place outside of the UK. Following the recent attack in Buffalo, the Home Office participated in international, multi-stakeholder debriefs to ensure lessons were learnt from this atrocious attack and improve the robustness of responses.

Although the attack was in the US, the attacker was heavily inspired by the Christchurch attacker and referred numerous times to the Christchurch attack itself within his manifesto and diary. Due to this, New Zealand was impacted domestically, and took appropriate action in response. The New Zealand Government contacted relevant communities, warned them of the potential news coverage and offered them support. These kinds of attacks can lead to the resurgence of past trauma for previous victims of attacks. During our engagement with the Christchurch Call Unit, it was explained that videos of both the Christchurch and Buffalo attacks were sent directly to victims of the Christchurch attacks with hateful messages.[90]

Other regulators, which implement take-down regimes, respond to these kinds of attacks mostly when complaints are raised to them. They then contact the platforms and request the removal of the content. For instance, even though the attack did not occur in Australia, the eSafety Commissioner identified multiple locations where this content could be found based on initial complaints about the availability of the video and the manifesto online. It then sent removal notices

[88] Discord. Discord transparency report, Jan to March 2022.
[89] BitChute [Twitter], 6 October 2022. Available here [accessed 6 October 2022]; Tech Against Terrorism. Announcing Tech Against Terrorism's Newest Member.
[90] Stuff, May 2022. March 15 survivors retraumatised by link to Buffalo attack livestream video.

to the websites, designated internet services or hosting service providers, no matter where they were located in the world.[91]

[91] The eSafety Commissioner has 'the authority to compel online service providers (social media services, relevant electronic services, designated internet services or hosting services) to remove seriously harmful content within 24 hours of receiving a formal notice'.

Implications for online safety

The Buffalo attack provides fresh insight into the ways in which threat actors can exploit the features and functionalities of online services, especially livestreaming, and how footage and other material linked to attacks can proliferate across online services. Interviewees highlighted four key challenges and choices facing companies seeking to prevent terrorist content as well as violent and hateful content from appearing on their services, which we discuss below.

User policies that balance safety and freedom of expression

Community guidelines, or terms of service, provide the starting point for platforms' enforcement of rules against terrorist, violent and hateful content. They guide both platforms and users in what types of content are allowed and what are not.

Our discussions with interviewees indicates that platforms face challenges in defining, locating, identifying, and removing content related to terrorism or violent or hateful content. There is currently no globally established or recognised definition of what this comprises, and the applicable criminal laws may vary by jurisdiction. In addition, judgements about what should be considered terrorist, violent or hateful content are often likely to depend on contextual factors which may not always be evident to platforms from the information available to them. It may be challenging for them to determine when to classify content as terrorist or hateful when making content moderation decisions.

One route taken by many platforms is to ban terrorist actors by reference to national or intergovernmental lists of proscribed or designated groups and individuals. It is easier, and more scalable, to block content from a particular source than to assess and act on individual items of content from any user.

However, platforms that only prohibit content from proscribed groups and individuals, and have no wider policies on violent and hateful content, would not have acted on the Buffalo footage, since the attacker was not affiliated with a proscribed group and was not a proscribed individual. These platforms would also not have acted on many other forms of terrorist, violent and hateful content, including certain content that could constitute a criminal offence in the UK. Further, an interviewee also suggested that there is a different approach to violent Islamist groups and violent far right groups, with Islamist groups more likely to be proscribed at present, and far-right groups typically under-represented. Due to this, platforms that prohibit content only from proscribed groups and individuals would not be acting against content from violent far-right groups that are not proscribed.[92]

Consequently, services may have internal lists of banned groups and individuals that go beyond national or international proscribed lists. Some services which allow livestreaming, prohibit livestreaming terrorist, violent and hateful content within their terms of service. Some services will also have more detailed supplementary policies relating to a wide range of content, including hate speech, incitement to violence, and disinformation/conspiracy theories that seek to radicalise users.

[92] European Eye on Radicalization, September 2022. Designation and Moderation of Online Terrorist Content.

For example, Meta's policies do not allow content that praises or supports terrorist attacks, or 'that praises, substantively supports or represents ideologies that promote hate, such as Nazism and white supremacy'.[93] As another example, YouTube's policies do not allow content 'intended to praise, promote or aid violent criminal organisations'.[94]

Services can set their own terms and conditions as long as they meet any specific conditions required by relevant laws and regulation. But this analysis shows that in doing so, they inevitably make trade-offs between protecting their users from potentially harmful content, and their ability to say what they want. For users to know what they may or may not share or encounter on platforms, they need clarity in the platforms' terms and conditions on how platforms define and tackle terrorist, violent and hateful content, and for the terms and conditions to be accessible to them.

Cross-industry collaboration: essential to counter terrorist content and activity online, but not a panacea

Cross-industry initiatives have sought to respond to terrorist use of the internet, to address the challenges of defining and identifying terrorist content and preventing its viral spread. These notably include GIFCT, the Christchurch Call, and the EU Internet Forum. While they all have a common goal, to adopt a multi-stakeholder approach (involving in particular civil society actors) and have their own interoperable crisis response protocols, the Buffalo incident has highlighted the challenges posed by their limited membership. This ultimately means participation in each, and thus the benefit of cross-industry collaboration, is currently limited to a relatively small group of online services, compared to the broader industry.

This was acknowledged at the Christchurch Call Leaders' Summit 2022, where heads of state and governments, leaders from online services and civil society committed 'to address the role of unmoderated and "alt-tech" services in disseminating terrorist and violent extremist content, day-to-day and in crisis'.[95] This further highlights the risk of exploitation of smaller and lesser-known services to spread terrorist and potentially radicalising content should be a focus of future work in this area.

TaT's focus and specialisation in supporting and helping smaller services improve their policies and processes seeks to address this challenge through its Knowledge Sharing Platform, TCAP, Mentorship Programme and Membership. While TaT does not entail sharing of information or collaboration between industry actors, it does have application across a broad section of industry. One area where it particularly adds value is through the TCAP's URL hashing database.

URLs are not currently hashed by GIFCT. The Buffalo incident showed how threat actors may upload content to small file-sharing platforms, which are not themselves members of GIFCT. URLs linking to those videos can then be distributed on bigger services, including GIFCT members. Currently these URLs cannot easily be identified and tracked across services. We understand GIFCT is currently working on incorporating TaT's URL hashing database into its existing image and video hash sharing

[93] Meta. Dangerous individuals and organisations [accessed 3 October 2022].
[94] YouTube. Violent criminal organisations policy [accessed 3 October 2022].
[95] Christchurch Call, September 2022. Co-Chair Statement Christchurch Call Leaders' Summit.

database, which would enable members to block these URLs following an incident in which the CIP was activated.[96]

As seen, industry collaboration continues to evolve.[97] It undoubtedly plays an important role in efforts to tackle terrorist threats online, although it may be that no single model can provide a comprehensive solution. Multiple approaches, addressing different services and different types of threat, may be needed.

Accessible user reporting tools and appropriately resourced content moderation

Many online platforms provide reporting and flagging mechanisms to allow users to identify harmful content. Previous Ofcom research has found that many users do not routinely use these tools or consider them effective[98]. Transparency reports also suggest that user notifications account for a small proportion of content takedowns, at least amongst the largest platforms.[99]

However, user reports may be relatively more important for livestreamed content, which is ephemeral, and less likely to be continuously monitored by automated tools. Users are more likely to flag harmful content if it is easy for them to do so. Systems that require individuals to be logged in before they can report and flag content, or force users to seek out additional information or take extra steps to make a report (such as sending an email) may deter and/or delay users from reporting content. This will tend to increase the risk that livestreamed content goes undetected for longer than necessary.

Reporting and flagging tools need to be backed up by appropriately resourced content moderation teams. User reports that go unaddressed may be one reason why users express scepticism about their value. While many platforms now have 24/7 moderation teams, some smaller platforms and start-ups may not have 24/7 coverage as they consider the costs and resource implications to be too great. However, periods where there is no human moderator presence on services may increase the risk that violative content is widely viewed or disseminated before being taken down.

There are risks that reporting and flagging mechanisms could be exploited maliciously to target non-violative content, and there are costs associated with implementing and improving accessible, real-time flagging tools. Nonetheless, as explained in our VSP guidance, accessible and user-friendly reporting tools are fundamental to effective user protection.[100]

Proactive management of risks associated with livestreaming

During the Buffalo attack, as well as many other terrorist and violent extremist incidents, the attacker exploited the particular characteristics of livestreaming: real-time, ephemeral and widely available. Though in proportion to the amount livestreaming is used in everyday life, terrorist

[96] GIFCT, June 2022. Debrief: CIP Activation, Buffalo, New York USA.
[97] For example, the Christchurch Call and GIFCT both conducted debriefs following the attack in Buffalo.
[98] Ofcom, June 2022. Online Nation Report 2022, page 71. A fifth of users reported or flagged potentially harmful content or behaviour they encountered.
[99] Meta. Transparency center: dangerous organizations: terrorism and organized hate [accessed 6 October 2022].
[100] See our VSP Guidance for providers on measures to protect users from harmful material, paragraph 4.59.

incidents are very rare, the harm caused by such a livestreamed incident can be significant. Further, as both the Christchurch and Buffalo incidents show, livestreamed footage can be subsequently widely disseminated or be a source of further radicalisation.

A platform seeking to reduce the risk of these incidents could consider introducing features that limit access to livestreaming. For example, YouTube and TikTok have age restrictions and only allow users to livestream once they exceed a minimum number of subscribers. Facebook only allows account holders to view livestreams.[101]

More broadly, platforms wishing to put in place effective risk management would make systematic efforts to consider the scope for certain functionalities or product features to be abused to cause harm and take steps to mitigate this risk. Some interviewees pointed out that while many companies have rigorous procedures to assess the privacy implications of new products and features, before launch and in use, it is less common for this to happen with respect to safety risks. The Buffalo attack has demonstrated the level of thought that threat actors place behind their actions when considering which platforms and functionalities to exploit. It is important that platforms wishing to prevent the upload of terrorist content put corresponding effort into making their services sufficiently robust against exploitation by these actors, and embed user safety considerations into the product and engineering design processes.

[101] TikTok. Going Live Video: A Walkthrough for Marketers [accessed 10 October 2022]; YouTube, Help. Create a live stream on mobile [accessed 4 October 2022]; Meta. Facebook Live [accessed 4 October 2022].

Conclusions and next steps

This work has helped build our understanding of the ways in which terrorist and violent extremist actors have used livestreaming to disseminate footage of their attacks, and the impact of industry responses and collaborative action. It has highlighted the particular challenges associated with livestreamed incidents:

- The use of multiple platforms by threat actors and their supporters, and the careful choice of different platforms to exploit their diverse vulnerabilities and maximise the chances of viral distribution of the footage;
- The difficulty of unambiguously identifying terrorist, violent extremist and hateful actors and content, with superficially similar content often posing very different risks and proliferating in different ways such as the Memphis incident.

In response to these challenges, the stakeholders we spoke to highlighted the importance for platforms to collaborate, to take account of the risks of adversarial use in product design and development, to have clear terms and conditions, effective user reporting tools and appropriately resourced content moderation teams.

We will take these insights into account in our continuing work towards the launch of the UK's online safety regulation. As appropriate, we will take account of this work in our future register of risks and risk profiles, risk assessment guidance for regulated providers and codes of practice, relating to illegal content. We currently plan to publish all these documents in draft form for consultation in Spring 2023, following passage of the Online Safety Bill.

We continue to work with UK-established VSPs to ensure they have appropriate measures in place to protect users from illegal terrorist, racist and xenophobic material; material likely to incite violence or hatred; and to assess any evidence we find of potential breaches of their duties and obligations under the VSP framework.

Finally, we continue to conduct research into the nature and prevalence of terrorist, violent and hateful content posing a risk to UK online users, which we will publish once complete.

A1. Stakeholder Meetings

We list below those stakeholders we spoke to for the purposes of the work underlying this report and who consented to be named. We have not listed those stakeholders with whom we met but have asked not to be named in this report.

STAKEHOLDERS	DESCRIPTION
NOTIFIED UK-ESTABLISHED VSPS	
BitChute	A peer-to-peer content sharing platform and associated services, with a focus on free expression.
TikTok UK	A short-form video sharing application.
Twitch	An interactive livestreaming service for different types of content, including gaming, entertainment and sports.
OTHER ONLINE SERVICES	
Discord	A voice, video and text communication social platform.
Google	Offers a range of services including Search, Maps, Gmail and Youtube
YouTube	A peer-to-peer video sharing platform.
CROSS-INDUSTRY INITIATIVES, REGULATORS AND GOVERNMENT BODIES	
Christchurch Call Unit, NZ Govt	A community of governments, online service providers, and civil society organisations acting together to eliminate terrorist and violent extremist content online.
GIFCT	An organisation that brings together the technology industry, government, civil society, and academia to foster collaboration and information-sharing to counter terrorist and violent extremist activity online.
Tech against Terrorism	An organisation that supports the tech industry to tackle terrorist exploitation of the internet, whilst respecting human rights.
Australian eSafety Commissioner	Australian independent statutory office with a range of regulatory functions and powers which

	aim to safeguard Australians at risk of online harms.
UK Home Office	UK government department that leads on a range of domestic issues including crime and counter-terrorism.
RESEARCH ORGANISATIONS	
Memetica	Memetica is a digital investigations group providing intelligence and risk advisory services on a variety of strategic issues relating to coordinated harassment, violent extremism, and disinformation.
Moonshot	Moonshot is a social impact company which builds solutions to understand and prevent online harm including violent extremism.
Institute for Strategic Dialogue (ISD)	An independent, non-profit organisation with expertise in extremism, hate and disinformation.
SUBJECT MATTER EXPERTS	
Brian Fishman	Co-Founder, Cinder and Former Director of Dangerous Organisations and Individuals at Facebook, expert in counter-terrorism.
Hany Farid	Professor at the University of California, expert in digital forensics.