POLICY BRIEF

Russian influence and interference measures following the 2017 UK terrorist attacks

Cardiff University Crime and Security Research Institute

POLICY BRIEF

The level of influence and interference by Russian-linked social media accounts trying to engineer social division in the UK is considerably more extensive than has been reported to date.

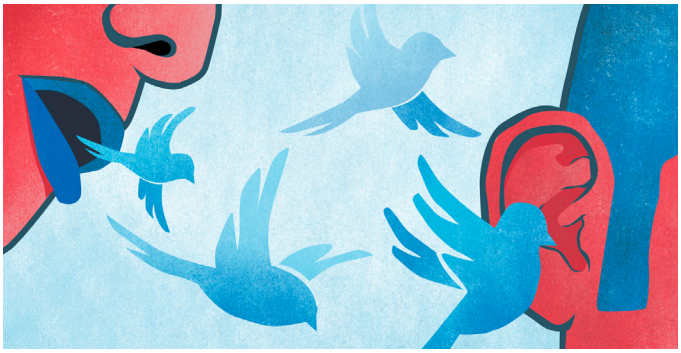
This brief details how independent analysis by CREST-funded Cardiff University researchers has identified systematic use of fake social media accounts, linked to Russia, amplifying the public impacts of four terrorist attacks that took place in the UK in 2017.

The evidence is that at least 47 different accounts were used to influence and interfere with public debate following all four attacks. Of these, 8 accounts were especially active, posting at least 475 Twitter messages across the 4 attacks, which were reposted in excess of 153,000 times.

We derived the identities of the Russian accounts from several open source information datasets, including releases via the US Congress investigations and the Russian magazine РБК. In addition to these 47 accounts, we have identified a number of others that possess similar 'signature profiles', but which have not been publicly identified as linked to the Internet Research Agency or similar Russian-linked units.

BACKGROUND

Claims of Russian agents engaging in operations aimed at influencing the trajectory of key events in Western countries have been well documented.



There are well-rehearsed and detailed allegations of them attempting to shape the American Presidential elections, the Brexit vote, and electoral processes in Europe. Further to which, the head of the UK National Cyber-Security Centre has also publicly stated that attacks have been committed against elements of the UK's critical national infrastructure, including utility companies and financial institutions. It has also been suggested that in the aftermath of the Westminster terrorist attack in March 2017, 'bots' connected to the Russian 'Internet Research Agency' helped to disseminate an internet meme encouraging anti-Islamic public sentiment.

The involvement of overseas agents in shaping the public impacts of terrorist attacks is more complex and troubling than the journalistic coverage of this story has implied. Specifically, there is evidence of such interventions:

- Involving a **greater volume of fake accounts** than has been documented;
- **Across four of the UK attacks** that took place earlier this year;

- Measures being targeted to influence opinions and actions simultaneously across **multiple positions on the ideological spectrum**;

- And, these activities are **not just being engaged by Russian units**, but also European and North American right-wing groups.

Consequently, we need a much 'richer', more sophisticated and comprehensive interpretation of what is happening, if we are to diagnose the intentions behind these actions and prescribe appropriate 'treatments'.

Terrorist attacks are designed as forms of communicative violence that send a message to 'terrorise, polarise and mobilise' different segments of the public audience. These kinds of public impacts are increasingly shaped by social media communications, reflecting the speed and scale with which such platforms can make information 'travel'.

Importantly, **what happens in the aftermath of such events has been relatively neglected by research and policy-development.** Far greater attention has been directed towards understanding terrorist motivations and intentions and how people are radicalised: trying to get 'upstream' insights about their moves towards violence in an effort to prevent its occurrence. Far fewer studies and policy initiatives have been directed 'downstream' towards **what works in mitigating the harms of attacks when they cannot be prevented.**

EVIDENCE AND INSIGHTS

Across four terrorist attacks in 2017 (Westminster / Manchester / London Bridge / Finsbury Park) we collected a **dataset of circa 30million datapoints** from various social media platforms. In processing these data, some anomalies were detected, which upon further investigation have been revealed to be associated with fake accounts.

A high proportion of the Russian accounts are positioned as 'breaking news' sites. There is robust research evidence that under conditions of crisis and conflict, people become more 'influenceable'. **Following the Manchester and London Bridge attacks, at least one account was sending inflammatory messages within 15 minutes.** In influence terms, responding rapidly to 'frame' the definition of the situation is important in being able to subtly shape how and what people think about something. There is an 'early mover advantage' to be accrued from getting in at the inception of an incident to try and sow seeds of antagonism and anxiety.

However, analysis suggests that these **breaking news sites were relatively ineffective** in terms of the numbers of 'impressions' they generated.

Far more effective and influential were 8 (out of 47) fake accounts, that generated a lot of information 'travel'. These accounts tended to be built around personal identities, clear ideological standpoints and were highly opinionated.

Following the four attacks, **a total of 475 original messages were posted from the identified Russian accounts** and these were **reposted in excess of 153,000 times.**

INCIDENT	NO. ORIGINAL MESSAGES FROM FAKE ACCOUNTS	NO. OF REPOSTS
Westminster	35	35,662
Manchester	293	55,581
London Bridge	140	57,322
Finsbury Park	7	4,871

Early communications interventions by these accounts attracted considerable support on social media. For instance, one tweet: *Another day, another Muslim terrorist attack. RETWEET if you think that Islam*

needs to be banned RIGHT NOW! Manches... (22 May 2017, 22:22) from an account adopting a right-wing, anti-Islam stance, sent less than one hour after the Manchester attack, was retweeted 3,606 times.

It is striking just how many followers some of these front accounts had. To take just three: @TEN_GOP (the right-wing, anti-Islam account mentioned above) had circa 127,000 followers on the 26 June 2017; @Crystal1Jonson (adopting a civil rights stance) had nearly 46,000 followers; and @SouthLoneStar (another with a right-wing stance) had almost 54,000.

An additional technique that those behind these accounts used to 'boost' their 'signal' was to deliberately target messages at celebrities and political figures with large follower bases. The idea being that this will help their ideas to 'travel' beyond the follower base associated with their own spoof identities and accounts. For example: *Hey @jk_rowling, why aren't you voicing your outrage at the Muslim terror attack on kids in #Manchester (23 May 2017, 20:19).* By engaging with these digital interactions, high profile figures have been responsible for spreading messages originating from false accounts, massively extending their reach. They are analogous to 'carriers' of a virus that unknowingly and unwittingly encourage its spread across a population.

Concurrently though, the spoof account owners were also aiming messages at 'thought communities' more sympathetic to and aligned with their online identities. For example, there are multiple instances of them '@-ing' Tommy Robinson, former leader of the English Defence League and Nigel Farage. The purpose being to try and **stir and amplify the emotions of these groups and those who follow them, who are already ideologically 'primed' for such messages to resonate.**

The quality of mimicry employed by those behind the false accounts is sometimes very convincing and hard to differentiate from the 'real' thing. This is an important aspect of the information dynamics overall, inasmuch as it is not just the spoof accounts pumping out divisive and ideologically freighted communications, they are also engaged in seeking to nudge the impacts and amplify the effects of more genuine messengers.

A number of the accounts involved were established relatively recently, but some have been in existence for a longer period of time. The first appears to have been set up in 2011, with a cluster also in the latter part of 2014 / early 2015.

As part of the analysis, **a number of additional accounts (circa 20) have been identified with similar profiles and signatures to those of confirmed Russian origin**, that have not yet been publicly linked.

The use of these accounts as 'sock puppets' was perhaps one of the most intriguing aspects of the techniques of influence on display. This involved two of the spoof accounts commenting on the same elements of the terrorist attacks, during roughly the same points in time, adopting opposing standpoints. For example, there was an infamous image of a Muslim woman on Westminster Bridge walking past a victim being treated, apparently ignoring them. This became an internet meme propagated by multiple far-right groups and individuals, with about 7,000 variations of it according to our dataset. In response to which the far right aligned @Ten_GOP tweeted: *She is being judged for her own actions & lack of sympathy. Would you just walk by? Or offer help?* Whereas, @Crystal1Johnson's narrative was: *so this is how a world with glasses of hate look like - poor woman, being judged only by her clothes.*

There were multiple further examples of the spoof accounts trying to propagate and project very different interpretations of the same events, consistent with their particular assumed identities.

COMMENTARY

In evaluating this analysis it is worth noting that, for independent researchers, it is difficult to substantiate

'beyond reasonable doubt' the provenance of these messages and the accounts from which they emanated, as they have apparently been deleted. Moreover, the fabrication work in disguising the real identities of the accounts, is skillful and they were well camouflaged.

Consequently, in piecing together the above narrative, we have been working with the 'digital traces and tails' of the online interactions in which these accounts previously featured. In many ways, it is an approach analogous with the methods of forensic science – a form of digital detective work. Indeed, there is a famous maxim that guides the work of forensic scientists that 'every contact leaves a trace'. The idea being that you can reconstruct an understanding of 'who did what to whom' by careful examination of the physical impressions that remain whenever two physical objects come into contact with each other. In a similar manner, our work has been working with the digital residues of a series of online interactions, that whilst no longer directly observable, nevertheless have left behind traces.

Consistent with the above, it is quite likely that there are additional accounts that we have not identified because they are more directly concerned with British and European issues. The fake accounts underpinning this analysis are more likely to have been concerned with American affairs.

POLICY IMPLICATIONS

In the wake of the 2017 terrorist attacks, platforms such as Twitter and Facebook were used to spread rumours, fake news and conspiracy theories to amplify and extend the impact and harm associated with the incident. Terrorist violence is fundamentally designed to 'terrorise, mobilise and polarise' its audiences, so if social media platforms are being 'weaponised' by third parties to amplify these effects, then they need to be required to urgently do something to mitigate this.

The evidence suggests a **systematic strategic political communications campaign being directed at the UK designed to amplify the public harms of terrorist attacks**.

Understanding that this is the contemporary landscape is crucial for both policy and practice development in this area. Both the head of the UK Security Service and the national lead for counter-terrorism policing have publicly stated, in recent months, that not all threats can be prevented through even the most assiduous investigation and intelligence work, and some attacks are likely to succeed in the future. However, as rehearsed above, most attention has been 'upstream' of the attack focused upon understanding and interpreting the motivations and intentions of potential violent extremists.

What has been **neglected is the downstream consequences and the potential for better managing and mitigating the harms associated with those plots that do get through. The implication is that we require a more sophisticated 'post-event prevent' stream to counter-terrorism policy.** Part of which, should focus upon rapidly establishing what counter-measures are effective in offsetting the impact of 'soft facts' propagated by overseas interests, as they seek to do the work of terrorist organisations by amplifying the capacity and capability of violent acts to mobilise and polarise Western citizens.

About this project

This research was funded via the Centre for Research and Evidence on Security Threats as part of a project focused on how 'soft facts' (rumours / fake news / conspiracy theories / propaganda) influence the aftermath of terrorist attacks.

Professor Martin Innes is Director of the Crime and Security Research Institute at Cardiff University.

For more information about the project, please visit the CREST website at www.crestresearch.ac.uk