ACICE Issue 07/23 (July)

ACICE Monthly Digest

A monthly roundup of significant news around the world

ADMM Cybersecurity and Information Centre of Excellence

Social Media

The Enigmatic World of Social Bots and Cyborgs

- In today's digital landscape, bots proliferate our social media platforms. Some of us may be familiar with conventional content polluters or spambots, but the latest iterations of bots are far more sophisticated and dangerous. These advanced entities are social bots and cyborgs, and they possess the capability to seamlessly blend into the online community, making them almost indistinguishable from genuine human users, and hence hard to identify. As such, they wield significant influence in the online space.

- Driven by powerful computer algorithms, social bots have an uncanny ability to generate large quantities of content that appear authentic, whilst interacting with social media users in ways that mirror human behaviour. For instance, they are able to build networks on social media like human users, including following other humans and bots, and tagging influential users and celebrities to enhance their credibility.

- Cyborgs, on the other hand, are hybrid human-bot accounts that display a combination of human-like actions and automated messages. While they may not be as swift in generating content as fully automated social bots, and require more maintenance and oversight, human intervention and moderation of content produced by cyborgs make them more convincing and harder to detect.

- Social bots and cyborgs can manipulate public opinion by disseminating fake news articles and engaging in convincing interactions with real users. A study by Nature Communications found that humans retweeted posts by bots and other humans at an equal rate for low-credibility content[1], which made it challenging to distinguish fake from real posts. Separately, in February 2023, Switzerland-based International Journal of Environmental Research and Public Health published study findings that tweets by human accounts were retweeted 1220.74 times on average, compared to tweets by social bots which were retweeted 1581.76 times, suggesting that information posted by social bots were more likely to be retweeted than genuine posts.

- These research findings were also reflected in real-life examples. During the height of the COVID-19 pandemic, social bots retweeted on the efficacy of hydroxychloroquine which led to an increased demand for the drug as a preventative measure against COVID-19. This is despite several randomised and controlled trials finding the drug ineffective. More recently in April 2023, the rollout of Twitter Blue strengthened the ability of social bots and cyborg accounts to impersonate credible sources and spread misinformation on the ongoing Russia-Ukraine conflict.

- The low barriers to entry for creating social bots and cyborgs contribute to their proliferation. According to the Canadian-based Journal of Medical Internet Research's findings in 2021, these bots and AI tools are readily available for purchase on the dark web, while numerous resources for building social bots can be found on free repositories like GitHub. This could pose dangerous implications, as nefarious actors seeking to use social bots and cyborgs for disinformation campaigns against certain countries, actors or corporations can do so with uninhibited access and ease. Such an open marketplace also gives rise to Disinformation-as-a-Service (DaaS) providers, allowing state and non-state actors to hire private contractors to wage damaging disinformation campaigns against one another.

- Another key advantage of social bots lies in their tireless work ethic. Unlike humans who need rest, social bots can operate around the clock, churning out content at a relentless pace that human users simply cannot match. This ability allows them to flood social media platforms with a deluge of automated messages, drowning out competing narratives by human users.

[1] Nature Communications defineed low-credibility content as content from low-credibility sources. Such sources are websites that have been identified by reputable third-party news and fact-checking organisations as routinely publishing various types of misinformation or disinformation, without adherence to professional standards of journalistic integrity.

- The fight against social bots and cyborgs has prompted innovative strategies for identifying and countering their influence. One such approach is the use of social honeypot accounts. Researchers can set up attractive, fake social media profiles to bait social bots and cyborgs, allowing the researchers to identify and analyse interactions with these suspicious accounts. US-based cybersecurity research company ZeroFOX put this into practice by using social honeypot accounts to lure social bots, and observing the social engineering attacks by 40,000 bots within a sandboxed environment. The research gave insights on how the attacks were carried out, the similarities and differences in various social bots' attacks, and analysed the attackers' possible motives.

- Another approach, proposed by the British Journal of Management in 2021, sought to apply Actor-Network Theory to differentiate between social bots and human users. The theory focuses on tracing the intricate network of connections and associations between actors (social bots or humans) and non-human entities (platforms, algorithms, etc) to identify trends and differences in the ways bots and humans act. For example, social bots tend to tag other bots rather than humans, given their limitations in engaging in sophisticated discourse, while lower-level bots usually mention or tag more sophisticated bots. Social bots also tend to share URLs and upload media content more frequently than human users, providing potential clues for detection.

A likely example of a Social Bot, with several giveaway characteristics (high tweet count, frequent sharing of URLs)

- These theories have also been applied by social media companies. For example, Facebook has employed a machine learning model called Deep Entity Classification (DEC), which employs principles of Actor-Network Theory to distinguish between genuine human users and automated bot accounts. DEC improves on existing detection processes by examining deeper features of a suspect account. It connects the account to its friends, groups and pages, and analyses various characteristics of these connected entities, such as age, number of friends, group members and page admins. By comprehensively assessing the behaviours and properties of these entities, DEC enabled Facebook to reduce the estimated volume of spam and spam accounts by 27%.

- With technological advancements, the arms race between bot-detection algorithms and bot creators will become more pronounced. Bot creators will learn more about the methods and theories used by bot-detection algorithms and will increasingly be able to design bots that circumvent these detection models. Moreover, the emergence of generative AI and large language models (LLMs) like ChatGPT will only improve social bots and cyborgs' abilities to generate more hyper-realistic texts, images, and audio that allow them to elude detection.

- Notwithstanding these, the fight against social bots and cyborgs must also be balanced with considerations for freedom of expression and avoiding over-stringent censorship. Policymakers and social media platforms need to tread carefully and cooperate to find the delicate equilibrium between safeguarding information and preserving an open online environment.

Terrorism

Death of ISIS-East Asia Province Emir

- On 14 June 2023, Philippine authorities announced that the Emir[2] for the Islamic State of Iraq and Syria (ISIS) in Southeast Asia, also known as ISIS-East Asia Province (ISEAP), Faharudin Hadji Satar aka Abu Zacariah, was killed in a military operation in Marawi, Lanao del Sur Province, in the Philippines.

- Several regional ISIS-aligned media units confirmed the death of the leader, and published eulogies glorifying his martyrdom.

- On 28 June 2023, Al-Fursan, a regional ISIS-aligned media unit, announced that ISEAP had identified the Bangsamoro Islamic Freedom Fighters (BIFF) leader Abu Toraipe as their new leader, claiming that he would "lead and establish ISIS in the region".

Terror Propaganda Targeting Southeast Asian Leaders

- On 20 June 2023, a pro-ISIS supporter shared a terror propaganda poster with a threatening message targeting several Southeast Asian leaders on the Element instant messaging platform.

- Key political leaders from Singapore, Indonesia and the Philippines were depicted in the poster against a backdrop of flames with the text "Hell is your end, O'enemies of Allah!". The four leaders shown (from right to left) were Halimah Yacob, the President of Singapore, Joko Widodo, the President of Indonesia, Ferdinand Marcos Jr., the President of the Philippines, and Ibrahim Murad, the Chief Minister of the Bangsamoro Autonomous Region in Muslim Mindanao (BARMM), the Philippines.

- In the same post, the recent death of the ISEAP Emir, Abu Zacariah was also mentioned, and the user encouraged others to perform jihad, which in Islam generally refers to meritorious struggle or effort. However, Islamic extremists commonly referred jihad to the fighting of enemies including the achieving of martyrdom.

[2] An Arabic term referring to a military commander or governor of a province in the Muslim middle-east region

Calls for Attacks in the Region

- On 16 June 2023, Indonesian ISIS supporters incited to attack Coldplay's concert in Jakarta slated to take place in November 2023. One of the supporters even shared an online tutorial on assembling a detonator which could be used to target and attack the concert.

- Following the burning of the Quran by Iraqi refugee Salwan Momika in Sweden on 29 June 2023, Indonesian ISIS supporters published posters calling for attacks on Salwan. There were also calls for Muslims residing in Europe to wage attacks.

- On 29 June 2023, in relation to the riots in France triggered by fatal police shooting of a teenager of North African descent, an Al-Qaeda unit celebrated the unrest and incited French protesters to resort to violence. They further encouraged the killing of French soldiers, and stated that it was "better" than having to migrate to a battlefield to fight against the enemies of Islam.

Annex

References

Social Media

1. Social Bots' Role in the COVID-19 Pandemic Discussion on Twitter (International Journal of Environmental Research and Public Health (Switzerland))
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9967279/

2. Social Bots and the Spread of Disinformation in Social Media: The Challenges of Artificial Intelligence (British Journal of Medicine (United Kingdom))
https://onlinelibrary.wiley.com/doi/10.1111/1467-8551.12554

3. Bots and Misinformation Spread on Social Media: Implications for COVID-19 (Journal of Medical Internet Research (Canada))
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8139392/

4. Understanding the Trolling Phenomenon (Journal of Information Warfare (Finland))
https://www.jstor.org/stable/26487554

5. The Spread of Low-Credibility Content by Social Bots (Nature Communications (United Kingdom))
https://www.nature.com/articles/s41467-018-06930-7

6. Facebook's New AI-Powered Moderation Tool Helps it Catch Billions of Fake Accounts (The Verge (United States))
https://www.theverge.com/2020/3/4/21164695/facebook-ai-moderation-dec-deep-entity-classification-fake-accounts-spam-scams

7. Honeypot Catches Social Engineering Scams on Social Media (Chief Security Officer (United States))
https://www.csoonline.com/article/560497/honeypot-catches-social-engineering-scams-on-social-media.html

8. How to Recognise Opinion Robots (Germany)
https://www.speiegel.de/netzwelt/web/social-bots-entlarven-so-erkennen-sie-meinungsroboter-a-1129539.html

Terrorism

1. Death of Islamic State Emir in Marawi Underscores Failure to Quash Remnants of 2017 Siege
https://pcij.org/article/10216/death-isis-emir-marawi-underscores-failure-quash-remnants-2017-siege

2. Islamic State Leader for Southeast Asia Killed in Marawi
https://www.benarnews.org/english/news/philippine/philippines-islamic-state-leader-killed-marawi-06142023035156.html

3. Top Regional ISIL Leader Killed in Philippines' Ruined Marawi
https://www.aljazeera.com/news/2023/6/15/top-regional-isil-leader-killed-in-philippines-ruined-marawi

4. Threats Will Not Backtrack Coldplay Concert in Jakarta: Tourism Minister
https://en.tempo.co/read/1726140/threats-will-not-backtrack-coldplay-concert-in-jakarta-tourism-minister

5. Explainer: Who is Salwan Momika, the Infamous Iraqi who Burnt the Quran in Sweden and Headed a Militia
https://www.newarab.com/news/who-salwan-momika-infamous-iraqi-who-burnt-quran

6. France Riots: Hundreds Arrested in Fourth Night of Unrest as 45,000 Police Deployed
https://www.theguardian.com/world/2023/jun/30/france-riots-more-than-400-arrested-as-police-officer-accused-of-shooting-teen-apologises

Contact Details

For any queries and/or clarifications, please contact ACICE at ACICE@defence.gov.sg

Prepared by:
ADMM Cybersecurity and Information Centre of Excellence