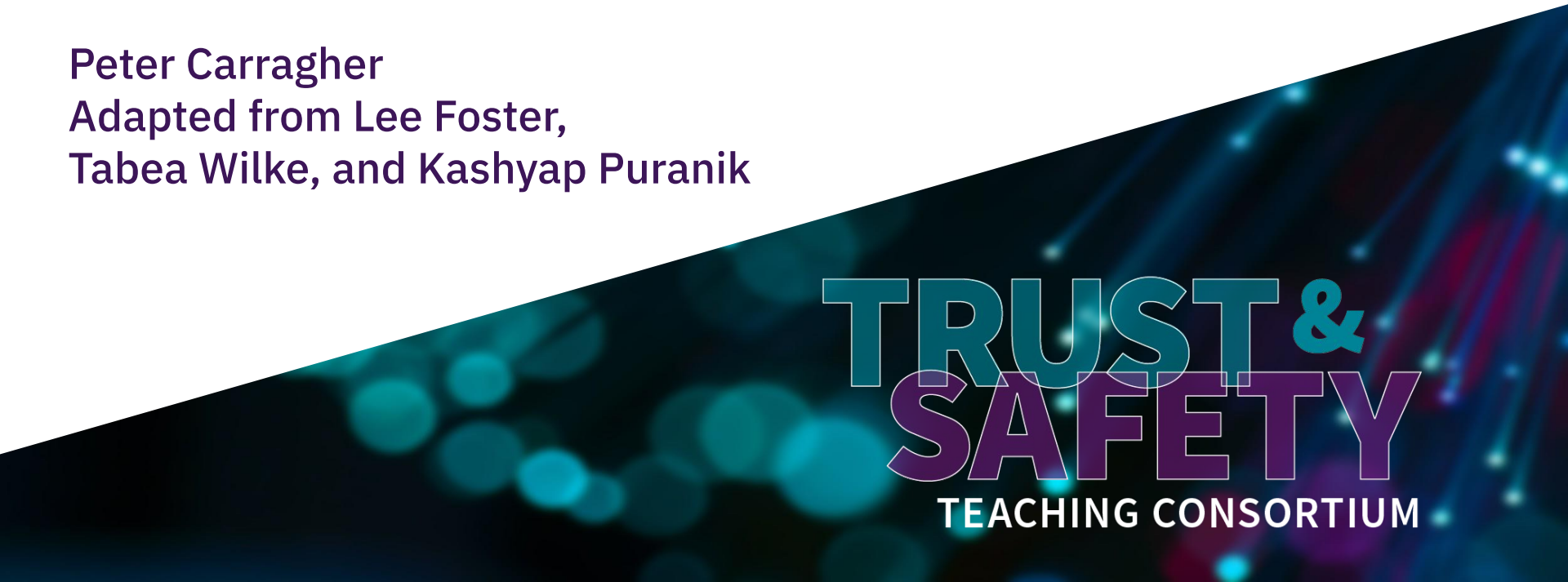
Authentication, Identity, and Platform Manipulation

Peter Carragher
Adapted from Lee Foster,
Tabea Wilke, and Kashyap Puranik



Learning Objectives

TRUST & SAFETY
TEACHING CONSORTIUM

Today we will:

- Develop a **conceptual understanding** of common social media manipulation tactics across key domains:
  - **Identity Abuse**
    - False Representation
    - Impersonation
    - Account Takeovers
    - Bulk account creation
  - **Synthetic Content**
    - AI-generated content and “deepfakes”
- Explore the real-world use of these tactics through **detailed case studies** of threat activity

TRUST & SAFETY
TEACHING CONSORTIUM

Identity Abuse

Authentication & Identity

TRUST & SAFETY TEACHING CONSORTIUM

- Not all platforms necessitate authentication or login for access. For instance, YouTube and Instagram allow browsing without requiring authentication.
- Authentication and digital identities serve several purposes
  - Products personalization and history tracking for users
  - User data protection from unauthorized access
  - Keeping online activity accountable
  - Rate-limiting logged out data access
    - Platforms like Twitter and Reddit have now started forcing login to limit scraping and data collection

Authentication & Identity

TRUST & SAFETY TEACHING CONSORTIUM

- **AAA** - Authentication, Authorization and Accounting
  - **Authentication** - Verifying user identity before granting access to a platform. Usually involves usernames and passwords. Cookies or tokens are presented by a client (Eg: browser) after authentication to identify the user
  - **Authorization** - Enforces granular access control and user privileges
  - **Accounting** - tracks user actions on a platform
- **Digital identity** - relationship between a human and their digital presence

Identity Abuse

Identity Abuse on social media can take multiple forms:

- **False Representation** (pretending to be something you are not, such as a journalist)
- **Impersonation** (actively adopting the identity of a real person)
- **Account Takeovers** (compromising and stealing the social media accounts of others)
- **Bulk account creation** (automated creation of accounts using bots)

Case Study: Distinguished Impersonator

TRUST & SAFETY TEACHING CONSORTIUM

TTPs

- Impersonation of U.S. Congressional candidates in run up to 2018 midterm elections
- False representation, posing as journalists to solicit interviews with Middle East policy experts and academics
- False representation, posing as Americans on social media and as citizens writing politically-charged letters and op-eds to local news entities in the U.S. and Israel.

Timeframe

- Approx. 2018-2020

Motivation/Attribution

- Support of Iranian political interests (low confidence)

Source

- FireEye/Mandiant: https://web.archive.org/web/20190529020105/https://www.fireeye.com/blog/threat-research/2019/05/social-media-network-impersonates-us-political-candidates-supports-iranian-interests.html

Case Study: Distinguished Impersonator

Accounts **impersonated** U.S. Congressional candidates on Twitter during the 2018 midterm elections:

- Republican candidates running for House of Representatives
- Plagiarized tweets from their legitimate accounts, interspersed with fabricated tweets

(Source: FireEye/Mandiant)

Case Study: Distinguished Impersonator

Accounts **falsely represented** as journalists and solicited interviews with various policy experts, politicians, activists, and academics focused on Middle East affairs

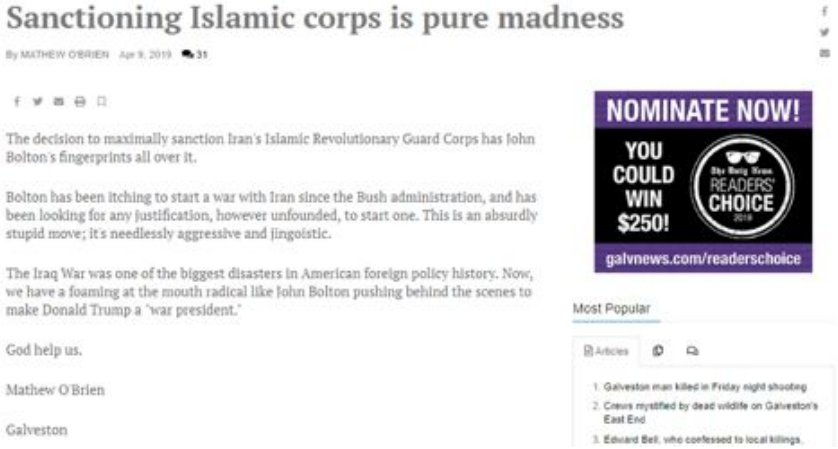
- Targets included US, UK, and Israeli Individuals
- Included video and audio interviews conducted over Skype
- Videos posted on Facebook, Twitter, and Instagram, some also posted to the Iran-based news outlet *Tehran Times*

(Source: FireEye/Mandiant)

Case Study: Distinguished Impersonator

Used **fictitious personas** to submit politically-themed letters, guest columns, and blog posts to local news entities in the US and Israel

- Personas claimed to be based in the locality of the respective news outlet (e.g. Baytown, Texas)
- Letters most often focused on the Middle East or US politics
- Geographies included Galveston TX, Baytown TX, Newport News VA, New York NY, among others



Sanctioning Islamic corps is pure madness

By MATHEW O'BRIEN Apr 9, 2019 31

The decision to maximally sanction Iran's Islamic Revolutionary Guard Corps has John Bolton's fingerprints all over it.

Bolton has been itching to start a war with Iran since the Bush administration, and has been looking for any justification, however unfounded, to start one. This is an absurdly stupid move; it's needlessly aggressive and jingoistic.

The Iraq War was one of the biggest disasters in American foreign policy history. Now, we have a foaming at the mouth radical like John Bolton pushing behind the scenes to make Donald Trump a "war president."

God help us.

Mathew O'Brien

Galveston

Most Popular

Articles

1. Galveston man killed in Friday night shooting
2. Cows mystified by dead wildlife on Galveston's East End
3. Edward Bell, who confessed to local killings,

(Source: FireEye/Mandiant)

Case Study: Spamouflage Dragon

TRUST & SAFETY TEACHING CONSORTIUM

Also known as: DRAGONBRIDGE

TTPs

- Fabricated social media personas (dozens of social media platforms)
- Use of synthetically generated profile pictures
- High production-value videos
- Attempts to motivate real-world protests

Timeframe

- Approx. 2018-Present

Attribution

- China (high confidence)

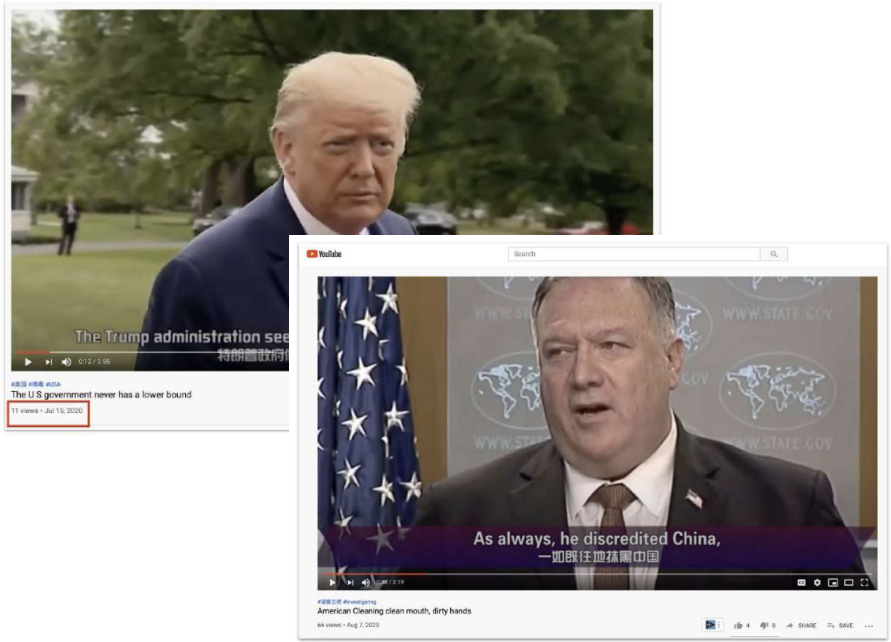
Sources

- https://graphika.com/reports/spamouflage
- https://graphika.com/reports/spamouflage-dragon-goes-to-america
- https://www.mandiant.com/resources/blog/pro-prc-influence-campaign-expands-dozens-social-media-platforms-websites-and-forums
- https://www.mandiant.com/resources/blog/dragonbridge-targets-rare-earths-mining-companies

Case Study: Spamouflage Dragon

Use **fabricated social media personas** to post and amplify large volumes of **politically-themed videos**

- Videos have so far spanned a wide range of topics, including:
  - Disparaging the 2019 Hong Kong protests
  - Pro-China propaganda
  - Negative commentary about US politics and domestic issues



(Source: Graphika)

Case Study: Spamouflage Dragon

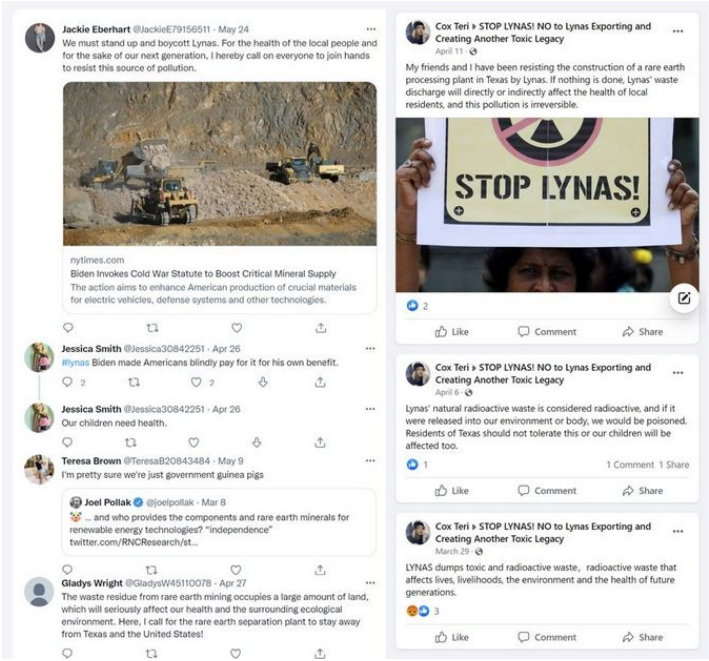
Often use **synthentically generated profile pictures** for social media personas

(Source: Graphika)

Case Study: Spamouflage Dragon

Have attempted to motivate **real-world protests** on various issues, including:

- Protests against Asian-American hate following racial attacks in the United States
- Protests against the construction of a rare earths processing plant in Texas



(Source: Mandiant)

Case Study: Ghostwriter/UNC1151

TRUST & SAFETY TEACHING CONSORTIUM

TTPs

- Fabricated journalist personas posting falsified stories
- Online impersonation of real journalists
- Hacking of legitimate news websites to post fabricated news articles
- Account takeovers of social media accounts of high-profile politicians in Eastern Europe

Timeframe

- Approx. 2016-Present

Motivation/Attribution

- Belarus (moderate-high confidence depending on component of activity)

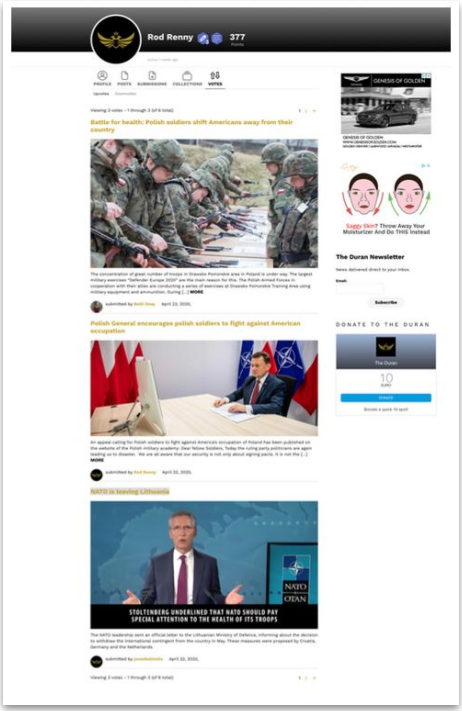
Sources

- https://www.mandiant.com/resources/blog/ghostwriter-influence-campaign
- https://www.mandiant.com/resources/blog/unc1151-linked-to-belarus-government

Case Study: Ghostwriter/UNC1151

Used **fabricated journalist personas and impersonated real journalists** to post fabricated news stories online

- Stories primarily focused on discrediting NATO and the presence of NATO troops in Eastern Europe
- Stories posted to third-party websites that allow publication of user-generated content
- Other stories published to hacked legitimate local news websites in Eastern Europe



(Source: Mandiant)

Case Study: Ghostwriter/UNC1151

Compromised legitimate social media accounts of high-profile politicians

- Compromised Twitter, Facebook, and Instagram accounts of Polish politicians to disseminate content
- Most likely compromised by first gaining access to their associated email accounts
- Primarily targeted members of political parties in Poland's United Right coalition
- Narratives intended to discredit the United Right and exacerbate domestic political divisions within Poland

(Source: Mandiant)

Account Takeover

TRUST & SAFETY TEACHING CONSORTIUM

Account takeovers provide abusive actors easy access to reputed and fully functional accounts. This can happen through many ways

- Account hijacking
  - Exposed password dumps and reused passwords
  - Password phishing/2FA phishing - users can be tricked into giving away passwords
  - Authorization phishing - users can be tricked into authorizing unwanted parties (Eg: OAuth phishing)
- Session hijacking
  - Cookie theft/session hijacking
  - Browser/device control through malware

Account Takeover

Preventing account takeover and potential negative impacts

- Prevention - Implementing Two-Factor Authentication (2FA) during login
  - **Pros:** Bolsters defense against unauthorized account access
  - **Cons:** May introduce inconvenience for certain users
- Prevention - Applying an additional layer of protection (2FA) for sensitive actions like account transfer, takeout, critical metadata changes
  - **Pros:** Minimizes user inconvenience
  - **Cons:** Increased implementation complexity
- Detection - Account takeover detection
  - **Pros:** Generally non-intrusive, barring occasional false alarms
  - **Cons:** Relies on the accuracy of detection systems

There’s a security vs usability tradeoff

Case Study: Physical security key as 2FA

TRUST & SAFETY TEACHING CONSORTIUM

- Google in 2018 July reported that it had not had any of its 85,000+ employees successfully phished on their work-related accounts since early 2017, when it began requiring all employees to use physical Security Keys in place of passwords and one-time codes. (Source: KrebsOnSecurity)
- SMS verification is vulnerable to man-in-the-middle attacks (MITM) and SIM swapping attacks while Authenticator apps are vulnerable to MITM.
- Not all second factors are equally effective.

Case Study: Cookie Theft

Cookie theft, also called 'pass-the-cookie attack,' is a method for hijacking sessions by stealing browser-stored session cookies. It's resurging as a threat, due to increased use of multi-factor authentication, making traditional attacks less effective. A sophisticated cookie theft campaign against YouTube creators reported in Oct 2021 by Google involved

- Social engineering creators with advertisement offers
- Fake software landing pages and social media accounts
- Delivering cookie theft malware
- Cryptocurrency scams and channel selling after compromise to monetize
- Hack-for-Hire attackers

Defenses involved hardening sensitive actions and login with 2FA as well as improved detection of malware websites and phishing emails.

(Source: Google blog)

Bulk Account Creation

TRUST & SAFETY TEACHING CONSORTIUM

- The mass creation of account offers malicious users greater potential to exploit a platform by generating harmful content or engaging in inappropriate behavior.
- Platforms often require verification of accounts through various means.
- There’s a security vs usability tradeoff
  - Requiring a state id verification makes a platform more secure and less vulnerable to abuse but hurts usability.
  - Relying on just phone verification makes platforms easy to use for both users and abusers alike.

Account Verification Mechanisms

Account verification mechanisms include-

- Text/phone verification
- Device verification
- Selfie verification
- State id verification

TRUST & SAFETY
TEACHING CONSORTIUM

Synthetic Content

Synthetic Content

TRUST & SAFETY TEACHING CONSORTIUM

Definition

- Synthetic content is not generated by humans, but with the help of automated and scalable tools
- Synthetic content can, but need not be, generated by Artificial Intelligence
- Synthetic content can take the form of:
  a) Text
  b) Images
  c) Audio
  d) Video

Synthetic Content

TRUST & SAFETY TEACHING CONSORTIUM

Relevance for trust & safety teams

Synthetic content is used for

1. **Impersonation**
   To impersonate or misrepresent individuals (such as public figures or heads of state) or events to generate false news or to shape the information environment

2. **Scaling information operations**
   To scale information operations on platforms like Facebook, Twitter, and YouTube, messaging channels like Telegram, publishing platforms like Substack or alternative news websites

Synthetic Content

TRUST & SAFETY TEACHING CONSORTIUM

1. Impersonation of public figures or heads of state

March 16, 2022: An information operation promoted a fabricated message alleging Ukraine’s surrender to Russia through an AI-generated deepfake video impersonating Ukraine's President Zelensky.

(Source: Google, Fog of war, p.40, https://services.google.com/fh/files/blogs/google_fog_of_war_research_report.pdf)

Synthetic Content

TRUST & SAFETY TEACHING CONSORTIUM

1. Impersonation of public figures or heads of state

May 2019: A video of Nancy Pelosi, former Speaker of the US House of Representatives, created the impression that she was drunk or seriously ill.

The video was amplified on the Facebook page “Politics WatchDog”. Within the first few hours, the video was watched more than two million times, shared more than 45,000 times, and had more than 23,000 comments.

Although the video was just slightly manipulated in its speed, it shaped the media agenda, the information environment, and the public perception of Nancy Pelosi.

(Source: Information Threats - Challenges for the European Information Space, p. 29, https://www.twincler.com/whitepaperinformationthreats/)

https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video

Synthetic Content

TRUST & SAFETY TEACHING CONSORTIUM

2. Scaling information operations

AI-generated profile images have been used to create profiles on social media networks. They support an operation, for example in amplifying narratives or commenting on posts, tweets, videos, or channels.

Source: https://public-assets.graphika.com/reports/graphika_report_operation_ffs_fake_face_storm.pdf

Synthetic Content

TRUST & SAFETY TEACHING CONSORTIUM

Impact of synthetic content for society

Even if synthetic content is low in quality, it is sufficient to shape the information environment and create confusion in sensitive situations for example during war, elections, a health crisis, natural disasters, terrorist attacks, or mass shootings.

Synthetic content can shape the outcome of major global events.

(Source: Information Threats - Challenges for the European Information Space, pp. 27, https://www.twincler.com/whitepaperinformationthreats/)

The first deepfake news anchor of the Chinese state broadcast station Xinhua.