The Big Book of Disinformation

Introduction

These are chapters from the Big Book of Disinformation Response, as used in the CTI League's Disinformation threat intelligence team for its Covid19 disinformation response. They've been modified - removing CTI-specific notes, adding in other notes that might help other similar teams. We've released them in the hope they can be useful to other teams running disinformation defence.

Chapter 0

These are chapters from the Big Book of Disinformation Response, as used in the CTI League's Disinformation threat intelligence team for its Covid19 disinformation response. They've been modified - removing CTI-specific notes, adding in other notes that might help other similar teams.

Contributors

We are community. We've been part of many communities on this journey, helped form communities and organisations, and have many people to acknowledge and thank for the work in these guides. Many people have been part of the BigBook efforts, and will be acknowledged here if they don't want to remain in the shadows.

CogSecCollab

The Cognitive Security Collaborative's mission is to bring together information security researchers, data scientists, and other subject-matter experts, in order to create and improve resources for the protection and defense of the cognitive domain; specifically in 2020, concentrating on building the tools and techniques that groups need to counter disinformation campaigns. It formed in January 2020 from the Credibility Coalition's Misinfosec Working Group, and the Misinfosec slack discussion channel, which built on work started in SOFWERX and the Hackers community.

The CTI League

The CTI League is a community of cyber threat intelligence experts, incident responders and industry experts working to neutralize all cyber threats looking to exploit the Covid19 pandemic. It formed in April 2020, and identifies, analyzes and neutralizes all threats but at this most sensitive time is prioritizing front-line medical resources and critical infrastructure.

The CTI disinformation team is tasked with finding coordinated inauthentic activities (“disinformation campaigns”), and the objects and people attached to them, and using

known ways (some of which we have to invent ourselves) to expose, disrupt or stop their operations.

The CTI League's disinformation team is embedded within the CTI League, and tracks disinformation using similar tools and techniques to the rest of information security, but there are some things that we do a little differently. Which is how we ended up writing a book...

Helpful info

Glossary

We all come from different disciplines: words like "campaign" have different meanings to a military, an adtech or a tech person (and if you're all three, you get to fight about definitions with yourself). There are also committees dedicated to defining what words like "disinformation" and "misinformation" mean, and the differences between them.

We ain't got time for that here. This glossary is our latest best effort at definitions for some of the words we use a lot between us, and what we (mostly) think we mean when we say them.

- Cognitive Security: The top layer of security, alongside Physical-security and Cyber-security. The art and practice of protecting against hacks that exploit cognitive weaknesses, especially cognitive hacks that are online and/or in large numbers of people. One of the reasons the MisinfoSec crowd started talking about Cognitive Security (including rebranding as the CogSecCollab) in 2020 is a belief that, in order to deal with things like disinformation, we need to focus on the thing we're protecting. That means working on reducing disinformation, but also on boosting good information when we see it.
- Misinformation: false content, where that content could be text, images, video, voice etc. Misinformation does not have to be deliberately generated (e.g. my mother might forget my favorite colour)
- Disinformation: deliberate attempt to deceive online. There is usually intent to deceive with disinformation, and the content itself might be true, but in a deceptive context (e.g.

fake users, fake groups, mislabelled images, doctored videos etc). Claire Wardle's work on the differences between misinformation and disinformation is still some of the best.
- Campaign: Campaigns are long-term efforts to change or confuse populations.
- Incident: Incidents are coordinated inauthentic activity that are carried out as part of a campaign. The “coordinated” implies either an instigator of some form with motives (geopolitics, money, ideology, attention, etc.) or some form of collective deliberate behaviour around it, like flooding a hashtag. That activity usually lasts for a short period of time because the narratives, artefacts, and other aspects can be picked up and continued by people who aren’t driving an incident - and this is often part of an incident or campaign’s goals.
- Narrative: Narratives are the “stories” that are being used to change minds, confuse people etc. Narratives are part of incidents - each incident might have multiple narratives involved, or just one, but there’s usually an identifiable narrative somewhere in there, that you can use to see if there are related incidents already tracked or dealt with etc. The other thing about narratives is that they, like incidents, have lifetimes. Some narratives appear as a result of a world or local event (or upcoming or anticipated event), and are only useful whilst that event is in peoples’ minds.
- Artefact: Artefacts are the objects that you can 'see' connected to a disinformation incident or campaign. They're the text, images, videos, user accounts, groups, hashtags etc that you use to get a picture of an incident or campaign.

Other terms related to this work:

- Astroturfing: creating a fake grassroots movement with an obfuscated sponsor or orchestrating group

Styles and formats

- We use ISO8601 format for dates where possible: yyyy-dd-mm (see https://www.w3.org/QA/Tips/iso-date)
- When referencing specific times related to incidents, explicitly declare the timezone or use UTC

Other places to look for information

There's a lot to learn about disinformation, misinformation, and how they fit into cognitive security / infosec in general - there's a separate [BigBook of Cognitive Security] being written about that. This BigBook is the practical one.

We've added lists at the end of this document (References), to books and papers about disinformation, to other teams doing this, sources of data, tools etc. And CogSecCollab is also collecting information in its documentation repo, which was used to seed this BigBook.

Chapter 1

What we're here to do

As the CTI Disinfo ReadMe says, "We're here to find, analyse, and coordinate responses to Covid19 disinformation incidents as they happen, where our specialist skills and connections are useful. We find and track new disinformation incidents, work out ways to mitigate or stop disinformation incidents and get information to the people who can do that."

1.5. How Disinformation fits into the Infosec Threat Response team

1.5.1. Activities

Reading through the CTI League handbook, the league stresses "Our members prioritize efforts on helping hospitals and healthcare facilities protect their infrastructures during the pandemic and creating an efficient channel to supply these services". The disinformation team should do this too.

It lists services as:

1. Neutralize malicious activities in the cyber domain with takedown, triaging, and escalation relevant information for sectors under threats.
2. Prevent attacks by supplying reliable, actionable information (IoCs, vulnerabilities, compromised sensitive information and vulnerabilities alerting).
3. Support the medical sector and other relevant sectors with services such as incident response and technical support.
4. Act as clearinghouse for data, a connection network and a platform for facilitating those connections

There are disinformation equivalents to these:

1. Neutralise: This is the disinformation takedown, triage and escalation work listed under disinformation incident response below.
2. Prevent: This is work that we could be doing - collating and supplying disinformation IoCs and vulnerabilities to the organisations, especially the health organisations, that

we work with. For example, if we identify that a “Reopen $STATE” campaign is attempting to organize another “Operation Gridlock” style incident, we can alert state, city, and county officials, as well as any hospitals in the target area.

3. Support: We’ve seen few direct cognitive security attacks on medical facilities so far. We have seen attacks directed at high-profile medical individuals and general attacks. We can assess the possibility of direct attack, and ways to be ready for that. For example, we could prepare resources that could be used in countering campaigns that target COVID-19 field hospitals.
4. Clearinghouse: We have connections established, but haven’t built ourselves as a clearinghouse yet. We could. We could also coordinate this work with those who are focusing on response and countercampaigns (the “elves” who fight the “trolls”).

For the neutralisation part, the league lists as examples:

- Infrastructures used by a threat actor that is exploiting the pandemic – malicious command and control server / DDoS servers / domains / IPs / etc.
- Exploiting legitimate services (such as open port in a legitimate website or compromised website used by hackers) and relevant to our stakeholders can be used to deploy attacks

The disinformation equivalents here would include:

- Hashtags, groups, networks, botnets, information routes, etc used by disinformation actor groups to create and run incidents. We can map several of these ahead of time, monitor them for new events forming (e.g. qanon checkins etc), and also file abuse complaints to registrars etc, notify companies hosting botnets and command and control accounts etc.
- Medical events (e.g. vaccination rollouts) that we know will trigger disinformation incidents

For prevention and support, the league lists examples:

- Alerting about vulnerabilities / compromised information and infrastructure to our stakeholders
- Creating a database of malicious indicators of compromise for blocking (via both MISP and GitHub repository)

- Alerting about trends and uneventful events regarding the pandemic in the cyber domain
- Creating a database of hunting queries for alerting systems.
- Create a safe and secure infrastructure for CTI League activities
- Create reports dedicated for the stakeholders and update them about ongoing trends of attack vectors regarding their organizations, such as significant information from underground-based platforms (darknet).

This is more detailed work, but as we track more incidents and become more familiar with the methods and tools used by incident creators, some measure of prevention activities become possible.

Team Structure and Communications

1.5.2. Channels and Bots

We have potential inputs, outputs and help across other channels, beyond our own channels.

- Data channels are useful for streaming supplementary input data that we don't want flooding the main human channels
- User channels are useful for finding us the people and places we need to get assistance, to report to (e.g. to find a specific Twitter group representative), to request takedowns etc.
- External team channels are other teams (e.g. darknet) who work alongside us, sometimes on the same artefacts
- Output channels stream clean outputs from teams

1.1 Coming in to help

The main work of the disinformation team is incident tracking and response. Live incidents are listed in theHive, and new ones are flagged in our slack channels as they’re added. We have 5 subteams supporting this:

- Incident management
- Tech
- Outreach
- Process and training
- People

The Triage team handles sensitive topics and content: this is a high-trust team, so we vet everyone who joins, starting with [filling out the disinformation team survey].

Getting help:

- When in doubt, ask a team lead: they can be reached in Slack with @disinfo-leads (we set this up so you could always get a lead). Otherwise, checking social media to see if a new incident is brewing is a never-ending job.
- For all things CTI Disinformation, start at the [Team Readme].

Team Leads

If the team is large enough, and running fast responses, a lead team is a good idea. Disinformation team leads also have a ReadMe and a BigBook for the things they need to keep a response ticking along. This is about responsibility: each lead is responsible for getting the thing done, not necessarily for doing it themself - a concept often forgotten in the heat of back-to-back responses. One way to allocate responsibilities across leads is:

- Overall lead: makes sure the disinformation team works smoothly and produces value. The lead keeps all the people and pieces in the whole team working well together: guides and supports all the team leads, arranging resources etc as needed, coordinates team activities, and tracks and logs team activities, keeping an eye on overall team health.
- People lead: Makes sure disinfo has the people it needs to do its job, and ensures there are routes to become a vetted (e.g. triage) disinformation team member. Arranges onboarding (newbie training, buddying etc) and vetting for prospective team members, maintains inventory of team skills available and needed, spots potential team trouble (e.g. troll breaches and other incursions), and offboards accounts if needed.

- Incident lead: Runs the disinformation incident response. Prioritises incidents, e.g decides which alerts to respond to, and which incidents to concentrate effort on. Finds and maintains effort (alerting, collection, analysis, mitigation, cleanup) on incidents, and decides when and how to hand off or close down incidents.
- Tech lead: Makes sure disinfo has all the tech it needs to do its job, and keeps that tech running. Build tools as needed. Finds and guides development talent as needed and appropriate for the disinformation team. Ensures tech builds are documented and repeatable / maintainable. If appropriate, this role might be accompanied by a research or data lead.
- Process and training lead: Maintains the processes and team skills needed for disinformation team to do its job. Maintains manuals (e.g. the BigBook of Disinformation Response). Manages team training: makes sure there *is* team training at least once a week, and that the people running it have the resources they need.
- Outreach lead: Maintains connections between disinformation and other connected teams. Manages alert and data sharing with other teams. Maintains connections to teams feeding data into disinformation, users of disinformation team outputs, and to sister teams whose tech/needs etc overlap with the disinformation team.

Chapter 2

if you're going to work on disinformation, you'll need to keep yourself safe. Some of the things you need to think about:

- Disinformation can be distressing material. It's not just the hate speech and *really* bad images that you know are difficult to look at - it's also difficult to spend day after day reading material designed to change beliefs, wear people down etc. Be aware of your mental health, and take steps to stay healthy (this btw is why we think automating as many processes as make sense is good - it stops people from having to interact so much with all the raw material).
- Disinformation actors aren't always nice people. Operational security (opsec: protecting things like your identity) is important
- You might also want to keep your disinformation work separated from your dayjob. Opsec can help here too.

Your Mental Health

Disinformation includes difficult material - it’s often designed to increase emotions like fear, hatred, disgust, to form in-groups and out-groups with hate speech and images that can be difficult to view, especially if they’re of a group you’re part of or feel strongly about. Even those of us who’ve been handling this material for years still get affected (that’s the point of it), so we all need to look after ourselves.

Some basics:

- Pace yourself if you’re going through difficult material.
    - Take regular breaks. Don’t spend more than an hour at a time reading through material.
    - If you can, arrange to be interrupted. It’s easy to get into a spiral with difficult material, and find yourself hours later still digging through it. Having an alarm, or a scheduled call from a friend, or the dog pestering you for its walk etc at the end of a session can stop this happening

    - If you can, go through material with a ‘buddy’ - pair up with someone online, preferably with a video or audio channel, and talk through what you’re doing with them.
    - Chocolate helps. We have no idea why.
    - If you start feeling wibbley, stop. There is no shame in this. Nobody in this team will ever judge you for taking a day, a week, two months off to look after yourself, or even shifting focus forever. Your mental health is important, and we will still be here when you’re ready.
- If you can avoid touching or reading material, do so. That means that, where we can, we automate. If we have 50 copies of the same image, we only need to view one copy, and if it’s a difficult image, not everyone on the team needs to see it.
    - If you have to share images / text in channels, put them in threads below content warnings, so people can choose whether to view them or not.
    - Automate feeds: if we have 50 copies of a message or image, only show 1 copy to the humans.
- Make disinformation something you “go to”. Right now, we’re surrounded by “the infodemic”. Friends are talking about it, feeds are everywhere, your great uncle is probably selling you the latest conspiracy theory. We’re also seeing most people in our lives online. Your life needs to include puppies and kittens, not being swamped by batshit crazy disinformation…(See [Basic OpSec for our Team] section below)
    - Don’t use your main social media accounts to follow disinformation. You don’t need more of that in your life. Pull the data you need using APIs; set up dedicated accounts to do the follows; ask the team if someone’s already following the accounts or groups you need data on.
    - Incognito mode. Nobody needs their ad feed full of Qanon t-shirts and bleach cures.
    - We won’t always be passive, so having some active accounts could be useful too...

1.3 Basic OpSec for our team

1.3.1 Key concepts

- Security. It’s a process. Tools help you execute the process.
- Compartmentation: separate your personal life from your work life.

- Persona: your spy disguise for research. A fleshed out human being that has details.
- Step 0: Lock your shit down.
- Goal: Impact containment. If you use compartmentation and a persona and everything goes wrong, all that gets compromised is the persona.

1.3.2 Process

OPSEC is a process, not a set of rules or tools. By continually following the process the user should remain in a state of security. The security you get is from following the process, not using tools.

1.3.2.1 Threat modeling for humans

EFF's Surveillance Self-Defense guide has a great introduction to threat modeling. In general, think about your 1-3 biggest threats -- in our case, revealing your real identity -- and consider the following:

1. What am I protecting?
2. From whom?
    1. What are they capable of doing?
    2. What's the worst that can happen to me?
3. How am I protecting myself and my info? (mitigate against them)

Once you've assessed your threat model, it's important to put it into action. Don't just sit there – do it!

1.3.3 compartmentalization: Engineering to make mistakes difficult.

An important part of operational security is implementing compartmentalization to limit the damage of any one penetration or compromise. compartmentalization is the separation of information, including people and activities, into discrete cells. These cells must have no interaction, access, or knowledge of each other. This is sometimes referred to as impact containment.

By compartmenting your operations, the control center over your accounts, and the information available from any single persona source, you are limiting the impact of a compromise. Without proper compartmentalization, attackers are able to leverage

information from one compromised account to access another related account. Increasing privileges and traversing across the persona's exposed and interlinked account control centers.

The strength of this compartmentalization is directly proportional to how strong your compartment walls are, and how well you maintain them. This takes discipline. But it isn't impossible.

1.3.4 Foundations: Personal Security

1.3.4.1 (Step 0) Baseline Security

Before you do anything else...

Secure yourself. Harden your personal environment.

- Implement unique, strong passwords everywhere
- Enable multi-factor authentication (2FA or MFA) on everything.
- Lock down privacy settings on your social media.
- Minimise your attack surface and exposure to retaliation if everything goes wrong.

Additional reading:

Security Guidelines for Congressional Campaigns

EFF's Surveillance Self-Defense Guide

1.3.5 Foundation: Work environment

1.3.5.1 Compartmentalization

No matter how good people get at hacking, they still have to obey the rules of physics.

Machines: Don't use your personal computer. Use dedicated equipment.

- At a minimum, use a Virtualbox VM.
- Better: use a separate, dedicated computer.

- Don't trust your brain to be perfect -- configure your computers differently so you have visual cues.
    - Use separate wallpapers and themes
    - Use separate browsers for separate tasks.
    - If you use dark mode on your personal computer or VM, set up light mode on your research computer or VM

Use a VPN: VPNs tunnel your internet traffic to make it look like you're in a different physical location. Use a paid product; if you're not paying a subscription for your VPN, the provider is collecting all of your traffic and selling the data.

If you're not sure which one, try ProtonVPN or Private Internet Access.

1.3.5.2 Cover: Your Persona

Once you've created your compartmented workspace, it's time to create a persona.

You're not trying to beat the NSA; you're trying to avoid being doxxed by trolls on 4chan. While it can be easy to go down a rabbit hole on this, you likely don't need a lot of backstory. With that in mind use a site like fakenamegenerator.com to create a persona.

Your persona should include at least:

- Name
- Email
- Phone number (non-VOIP burner works best if signing up for accounts)
- Account usernames and passwords
- Address
- Birthdate

Keep this info in a text file and leave it on the desktop of your working machine.

1.3.6 Work recipes (if this, then that)

Need to get people to explain the process of what they're doing, so we can build out the relevant recipes

- OSINT Research

Always start with Step 0: Baseline Security

This is intended as a quick and dirty guide to considering your Operational Security (OpSec). Consider this a starter guide or Level 0. There is a baseline for security to protect yourself, your fellow researchers, and the project. Obviously your approach to OpSec is going to depend on your threat model. Given the current context I'm going to skip an in depth discussion of physical security in favor of other topics.

1.3.7 OpSec Appendices

The starting point for building security is to limit the potential impact of a compromise. To contain the damage from a compromise use the principle of compartmentalization. Build a strong secure compartment to use for all your work and ensure their is no taint or contamination from inside the compartment back to you.

1.3.7.1 Threat Modeling

(from Lorenzo Franceschi-Bicchierai's What is Threat Modeling?)

"The first step to online security is figuring out what you're trying to protect, and who you're up against.

To help you figure out your threat model, consider these five questions:

1. What do you want to protect?
2. Who do you want to protect it from?
3. How likely is it that you will need to protect it?
4. How bad are the consequences if you fail?
5. How much trouble are you willing to go through in order to try to prevent those consequences?

By answering those questions, and figuring what solutions and tools you want to adopt based on them, you will come up with a threat model that works for you.

Overestimating your threat can be a problem too: if you start using obscure custom operating systems, virtual machines, or anything else technical when it's really not necessary (or you don't know how to use it), you're probably wasting your time and might be putting yourself at risk. At best, even the most simple tasks might take a while longer; in a worst-case scenario, you might be lulling yourself into a false sense of security with services and hardware that you don't need, while overlooking what actually matters to you and the actual threats you might be facing."

1.3.7.2 Physical Security Basics

- Cover your webcam to prevent unauthorized access to your camera.
- Lock and password protect computer
- Enable full disk encryption
- Optional: If you're concerned about unauthorized access to your microphone, you can use a mic block. Here is one example.

1.3.7.3 Passwords

Weak passwords and password recycling are the easiest ways to have your accounts pwned

- Haveibeenpwned: Check if your email account has been compromised in a data breach.
- Most password managers will alert you if your password has appeared in a data breach.

1.3.7.4 Password Managers

Password managers are the easiest way to create, store, and implement secure passwords for all your accounts.

Decision Point: Local or cloud-based password manager.

- Local: more secure, less efficient, harder to maintain, easier to lose everything if you forget to back up or lose access to your local version
- Cloud-based: easier to use, accessible anywhere, more efficient, less secure

Some options:

- 1password (cloud-based)
- LastPass (cloud-based)
- Dashlane (cloud-based)
- KeepassXC (local)

1.3.7.5 Two-Factor Authentication (2FA)

Two-Factor Authentication requires the user to provide an additional form of verification beyond just their password (Something you have + something you know). After having a strong unique password for each account, adding 2FA to an account is the highest leverage way to secure your account against unauthorized access.

- Two-Factor Authentication Handout from the EFF
- Twofactorauth.org: List of websites and whether or not they support 2FA.

Decision Point: Method for 2FA

- Text message (SMS): Easiest to get users to adopt, least secure, especially in our context. If you use it, best to use a burner VOIP number.
- Soft token (App-based): More secure than SMS. Examples include Google Authenticator and Authy.
- Hard token (Physical device): Most secure, harder to implement. Examples include Yubikey.

1.3.7.6 Using a VPN

A VPN is a program that routes all of your internet traffic through a different IP Address (like a tunnel). A VPN is one of the most effective ways to maintain anonymity online. Since VPN's basically route all your traffic like an ISP would, be sure you trust the provider. This is one of those things you should pay for, because if you're not paying for the product, you are the product. The VPN market is a racket; the review sites are a part of that. I've found thatoneprivacysite's reviews to be useful.

Here are some VPN options I've found helpful:

- ProtonVPN, by the same folks that make Protonmail
- Private Internet Access

Check that you're VPN is working properly by going to ipleak.net

Decision point: VPN on your network, on your device, or both

- On the network:
  - Pro: Filters all traffic from all devices on your network, not just web traffic or one device. If you lose VPN connection you can kill all internet access so nothing gets through without going through the VPN
  - Con: Longer and more complex setup and you need a dedicated device
- On your device:
  - Pro: Quicker and easier to get set up. Doesn't require any extra equipment.
  - Con: Only filters traffic from your one device and if it fails you may not realize immediately (unless it has a reliable killswitch). Also data your computer sends back to services on startup may get through before the VPN kicks in.

1.3.7.7 Web Browsers and Extensions

Decision Point: Which browser to use for general investigations

My browser of choice: Firefox

Essential Extensions

- Install Firefox Multi-Account Containers lets you separate your work, shopping or personal browsing without having to clear your history, log in and out, or use multiple browsers. Container tabs are like normal tabs except that the sites you visit will have access to a separate slice of the browser's storage. This means your site preferences, logged in sessions, and advertising tracking data won't carry over to the new container. Likewise, any browsing you do within the new container will not affect your logged in sessions, or tracking data of your other containers.
- Install Privacy Badger a browser add-on from the EFF that "stops advertisers and other third-party trackers from secretly tracking where you go and what pages you look at on the web.
- Install uBlock Origin, a wide-spectrum content blocker.
- Install HTTPS Everywhere, a browser extension from the EFF that encrypts your communications with many major websites, making your browsing more secure.

1.3.7.8 Burner Email and Phone numbers (pseudonymous identities)

In the process of doing investigations, you will likely find yourself in a position where you want to create burner accounts that allow you create pseudonymous personae. When possible, I create a full identity with name, email address, VOIP phone and text as well.

- Sudo: In terms of an easy to use pseudonymous identity, I've found that sudo is a great, easy to use option. It is a paid service, so that can be a barrier, but it allows you to create a personae and associate and isolate email, phone calls, text, web browsing and payment for each persona.

1.3.7.8.1 Burner Emails

Depending on your needs you may wish to create anonymous/pseudonymous emails. These are disposable temporary email addresses you can use. Many of these will get flagged by social media services as suspicious, so it's good to know about different options.

- 33mail Free option that might get flagged
- Protonmail: Free end-to-end encrypted email
- Gmail: quick and easy commercial option that will pass muster for most services. May have issue with this if you try to sign up for a bunch with the same phone number (which you shouldn't do anyway)

1.3.7.8.2 Burner Phone and phone numbers

There are tons of ways to get a free VOIP account. One challenge with VOIP numbers is that some services you'll want to use require a real phone number and won't accept VOIP for account registration.

- Free VOIP: Google Voice. You'll obviously need an associated Google and account and getting it requires providing a real phone number (major downside).
- Paid VOIP: Burner, Hushed, CoverMe
- Burner phones: Lots of different options including Tracfone where you can get a cheap phone and swap the SIM when needed.

1.3.7.9 Secure Communications

Use End-to-End Encryption (E2EE) wherever possible. E2EE is a system of communication where all data is encrypted in transit and at rest, meaning no one (including employees at

the company) has access to the data except the communicating users. This is the closest you're going to get to a completely private and secure way to communicate and store data.

1.3.7.9.1 Secure Messaging

End-to-End Encrypted messaging generally requires both users to be on the same service. This often means that the best service is the one with the most people you're trying to communicate with. Here are a few options:

- Signal is great and the How to Use Signal on iOS from the EFF is helpful. Popular among infosec, privacy enthusiasts, and journalists. One downside is that you have to tie the account to a real (non-VOIP) phone number.
- Whatsapp: Most popular E2EE messaging app. Built on the same encryption protocol as Signal. Major downside: owned by Facebook.
- iMessage: Incredibly popular. Only available to Apple users. E2EE breaks down depending on how you configure its relationship to iCloud for backing up messages.
- Others: Wire, Wickr, etc

1.3.7.9.2 Secure Email

End-to-End Encrypted email services:

- Protonmail
- Tutanota

1.3.7.9.3 Secure Ephemeral Communications:

- Firefox Send uses end-to-end encryption to keep your data secure from the moment you share to the moment your file is opened. It also offers security controls that you can set. You can choose when your file link expires, the number of downloads, and whether to add an optional password for an extra layer of security.
- CloakMy: quick, convenient and secure way to share sensitive information. Just copy your message in the box, set the recipient and your password (if you want to protect your message) and send it. The recipient will receive a secure link. If you select Auto Destruct as an expiration setting (by default), once the link is opened the message will be deleted. The message will be encrypted with a randomly generated key + your password if you chose one.

1.3.7.10 Social Engineering and Phishing

Phishing happens to everyone and it sucks. Here are a few ways to avoid getting phished.

- Urlscan.io allows even inexperienced users to investigate possibly malicious pages, such as phishing attempts or pages impersonating known brands.

A few other things to consider (which I hope we can expand upon later)

- Turn off location services on everything possible
- Locking down the setting on your social media accounts
- Removing yourself from people search sites (in case you get doxxed)
- Remove metadata from your photos before you post them
- 'This person does not exist' generates very convincing faces, again using machine learning. Reload the page to see another image. As the name suggests, these are not real people - the faces are generated entirely automatically. You can see artifacts, especially in the teeth, but this is still very close to perfect (and of course great for creating fake users).

Chapter 3

Defining Disinformation

We look at disinformation as an information security threat. It helps to define how we see the threat, its sources and its manifestations.

The short answer on defining disinformation is “don’t get hung up on definitions”. There are many definitions of disinformation, misinformation, malinformation, propaganda, influence etc., and standards working groups dedicated to defining the differences between them. If you’re doing practical disinformation response, try not to get sucked into that very large rabbithole: pick a working definition for yourself, work out what matters to your practical work, and focus on that.

For instance, the one we're using here comes from the Credibility Coalition's Misinfosec Working Group: "deliberate promotion of false, misleading or misattributed information. We focus on the creation, propagation and consumption of disinformation online. We are especially interested in disinformation designed to change beliefs in a large number of people".

That allows us to talk about

- intentionality (“deliberate promotion”),
- non-false information (“misleading or mis-attributed”),
- goals (“designed to change beliefs in a large number of people”) and
- mechanisms (“focus on creation, propagation, consumption of misinformation online”).

Within another organisation, we switched from trying to define misinformation in websites, to looking at signals of intent, e.g. did these sites contain hate speech, were they targetting specific groups etc. That moved the definitions from subjective, intangible and subject to bias (e.g. political sites are very difficult to flag as misinformation/not), to more objective tagging.

Disinformation isn't the same thing as misinformation.

- Misinformation is false content: untruths in text, faked images etc; and those might be unintentional, or not be part of a coordinated effort. Disinformation is false. It's intentional. It's at scale. And the falsehood might not be in the content - the content, or the original poster's intentions, might be clean, but the reuse, or amplification etc might be designed to create harm.
- A third category is Malinformation: information that's true, but usually private, and posted online to cause harm.

A good place to start if you want to dig into these definitions is Clare Wardle's 2017 work on Information Disorder.

Disinformation and Malinformation are examples of online harms, alongside things like ransomware, cyberbullying, etc. Always look for the harms, and the motivations for those harms.

Good introductions to disinformation

If you say you're working on disinformation, people around you will often quietly ask how they can help, and where they can get more information about it. Good introductions that you can show your mum and other people who ask include:

- The War on Pineapple: Understanding Foreign Interference in 5 Steps
- Bad News Game
- The Dark(er) Side of Media: Crash Course Media Literacy #10
- Web Literacy for Student Fact-Checkers – Simple Book Production

Although that doesn't cover everything we do, those references between them give a good introduction to what we're dealing with, and some of the things that everyone can do to help mitigate them.

1.1 Online Influence and its abuses

Let's look at the range of ways users and groups are influenced online (and offline via online means) - user experience, marketing and adtech, online political campaigns, astroturfing, online psyops, disinformation campaigns.

1.1.1 What People Do Online

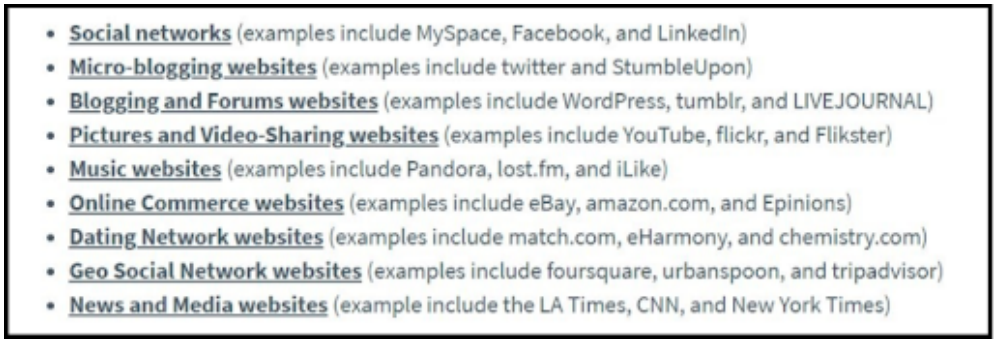


Figure: Chris Burgess, types of online interactions

The internet has changed a lot since the early days of ARPANET, JANET and bulletin boards. People still do the same things - sharing information and talking to each other - but the ability to do that isn't limited to the techies and companies who could pay for website designs, and the volume, variety and velocity of information and the people and organisations receiving it has increased to encompass (through localisation, phone apps etc) a large proportion of the world's population. Anyone can broadcast to almost anyone else almost instantly over a large number of specialised (shopping, music, dating, games, news, entertainment, research, etc) and general (social media, blogs, new websites etc) sites, through user-generated content like messages, posts and comments, and commercial content like videos, articles and pages.

1.1.2 With people comes value

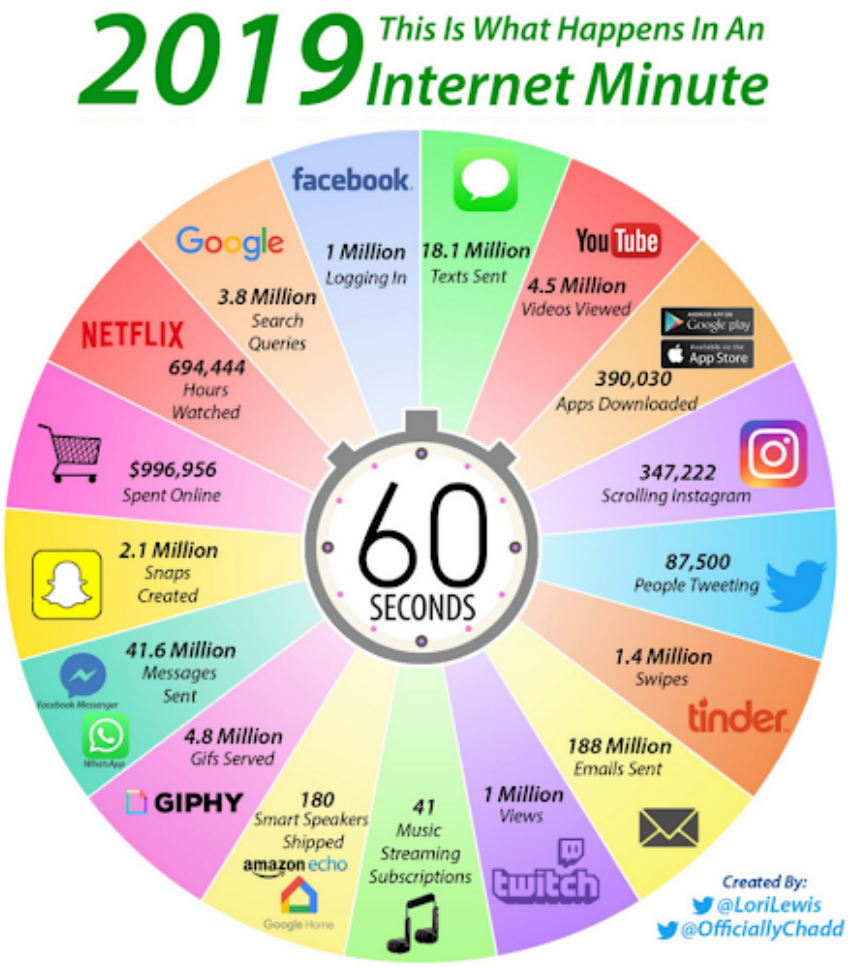


Figure: Internet Minute, 2019

If you create an online application or platform, there are several ways to make money:

- One-off payments (e.g. buying a t-shirt from an online vendor)
- Commissions (e.g. Amazon percentages on marketplace)
- Subscriptions (e.g. Spotify premium, AWS, New York Times etc)
- Online advertising (e.g. selling advert views, clicks, actions on your webpages or videos)

Money isn't the only commodity available where large numbers of people congregate. Other values include:

- Viewpoints. The stances that people take on issues like, for instance, who downed MH17.
- Belonging. Finding your community is much easier with billions of people online.

- Convening power. Sites like Eventbrite and Meetup help users build communities offline.
- Connections. Visibility builds relationships - whether this is with online dates, friends of friends or brands, products and influencers.
- Information.

1.1.3 With value comes targetting

Much of the online advertising industry is geared to optimising the high-speed auction between advertisers and online property owners (websites, videos, TV, internet-connected billboards etc), to get advertisers coverage whilst optimising the property owners' profits. What they're selling is users' views and actions. And what they optimise on is demographics (for individuals) and Know Your Customer (for businesses).

- Demographics: know your targets
- B2B: Know Your Customer

1.1.4 With value comes abuse

The difference between online marketing and disinformation campaigns is in intent. It's why we talk about "coordinated inauthentic activity", which focuses on the scale, the behaviour (you can do a good disinformation campaign with true content - e.g. almost any african-american focussed one) and the intent to deceive - where that intent is usually to do some form of harm, whether it's to shape a geopolitical narrative away from the country it's targetted at, or to widen divisions across society. Most disinformation campaigns look like marketing campaigns because that's where their roots are. The Internet Research Agency was a marketing team that was asked to do a side gig; many of the new disinformation farms in e.g. the Philippines are repurposed spam factories etc.

Disinformation as a Digital Harm

Disinformation is just one form of online abuse, amongst hate speech, spam, online bullying etc. These are often known collectively as "digital harms".

Reading

Internet history:

- An Internet History Timeline: From the 1960s to Now
- https://www.slideshare.net/debbylatina/internet-history-190741201

Internet size:

- We Are Social: Global digital report 2019

Abuses and counters

- I stumbled across a huge Airbnb scam that's taking over London
- Ethan Zuckerman course, "Fixing Social Media"
- There are no sharks swimming on a freeway in Houston
- Kate Starbird, 2016 Tracing Disinformation Trajectories from the 2010 Deepwater Horizon Oil Spill

---

Disinformation from the Creators' POV: Intent

We track disinformation incidents and persistent threats. Understanding what disinformation creators do can be improved by first understanding why they do it, and seeing how they might optimise against those goals.

Where disinformation comes from

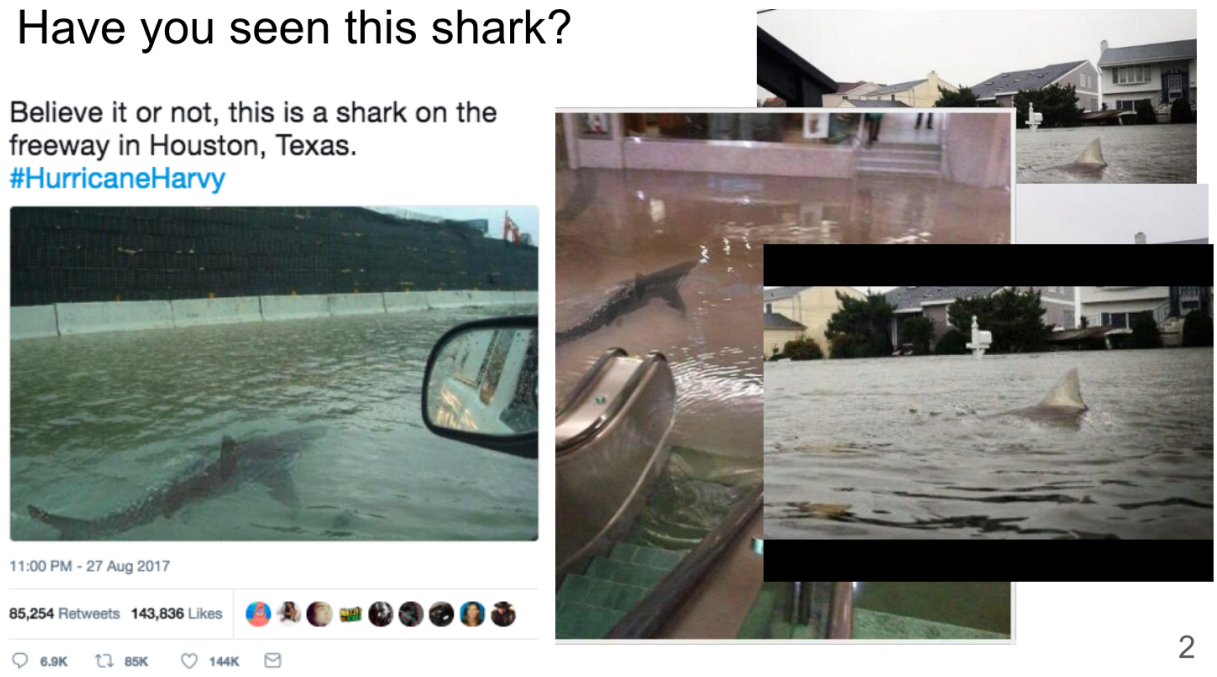
The short answer is that people produce disinformation for attention, power, money and political or geopolitical gain.

- Using disinformation for geopolitical gain is very much in the headlines. Countries use it to change opinions of themselves, their actions, and the state of areas they have interests in, and to weaken the population and environments of their potential opponents. Disinformation is cheaper than conventional warfare, with very few current downsides for a country willing to use it, can be outsourced to small teams and

individuals outside the country using or the subject of it, and done right will continue in the target country long after the creating team has moved on.

- Internal groups and organisations also use disinformation to gain power, often by emphasising ingroup/outgroup narratives to create strong groups of followers.
- Money is a popular motive,and even with other types of disinformation, there are often hucksters riding narratives and groups to make profits.
- Attention-seeking with online disinformation has been around a long time (e.g. the sharks in the street that appear online for most natural disasters, and satire and other LOLs); usually it's smaller-scale and driven short-term by individuals. Mostly, unless it's DDOSing a hashtag or area that's important (e.g. a crisis reporting hashtag) this type of disinformation gets lost in the noise.
- The social internet is driven by community: online discussion contains a lot of misinformation, including rumour, opinion, conspiracy theories, protests, extremists and combinations of them. This is humans being humans. We're not here to stop debate: disinformation tracking is about finding the coordinated inauthentic activities that potentially do harm.

1.2.3 Disinformation for Lols and Attention



The disaster shark. Every natural disaster, the same shark

Misinformation-for-fun going viral online has a long history. One classic example of this is the disaster shark pictures: in almost every natural disaster in the last decade, someone has posted a picture of the same shark as “sharks in the street”, “sharks in the subway” etc, and pushed it to go viral. Crisismappers see it and sigh, then ask the original poster to remove it please because it’s messing up the online response (crisismappers are typically social listening on disaster-related hashtags, looking for information they can add to a disaster situation picture and/or route to responders). Typically, people posting misinformation for fun are amenable to helping counter any ill effects from it, and are less likely to engage in counter-counter games.

T-shirt of fake tweet sent during Chile 2010 earthquake

Misinformation for attention also has a long history. Crisismappers have worked since 2010 to remove both for-profit (Ugg advertisements) and well-meaning (“thoughts and prayers”) spam from their feeds, but also to handle deliberate injections of what looks like real disaster-related information, both from individuals seeking attention (e.g the Chile t-

shirt tweet above), and from nationstates testing disinformation mechanisms (see Kate Starbird's analysis of the 2010 BP Oil Spill "tsunami warning" tweets on #oilspill). Generally, crisismappers triple-verify, e.g. don't post any information until we've received it from 3 sources and checked each of them out; for misinformation the process is to gently push back with a message (gentle humour can be good), and to reach out and ask the poster to remove it from social media.

1.2.4 Disinformation for Money

Money: It's grifting. Ways to make money from disinformation:

- get people to look at your website (cpm: $ for every thousand eyeballs), or click on something (cpc: $ for every click - a lot higher than cpm because it's a lot rarer), or do something like fill out a form (cpa - much much rarer usually); inadvertantly give you data that you can sell
- sell merchandise like t-shirts, videos, 'cures'; sell services (e.g. the covid5g guy selling books and on a speaking tour);
- sell disinformation services (e.g. like spam farms, but for disinfo, or creating deepfakes - but at about $2 per hour and 5-6 hours per fake that's not making much right now);
- sell or rent accounts (e.g. botnets - again, still relatively cheap)

In 2016, it was the "Macedonian Teens" (in practice, not really, but there were some villages that were centres of this), and US-native peeps who discovered that political outrage on the right side of the spectrum got more clicks. Basically: anger and fear sell

Now, there are a lot of antivax sites selling 'alternative cures' Also good bit of affiliate marketing - usually to its own network of sketchy sites.

1.2.5 Disinformation for GeoPolitics

Since 1648 (the end of the 30 Years War, where over 8 million people died), modern international discourse between nations has been based on Westphalian Sovereignty. This includes the principles:

Each nation has sovereignty over its own territory and domestic affairs No nation should interfere in another country's domestic affairs Each state is equal under international law

Nation states influence each other through the instruments of national power. These are resources available in pursuit of national objectives, usually referred to using the DIME model [74]:

Diplomatic: Diplomacy is a principal means of organizing coalitions and alliances, which may include states and non-state entities, as partners, allies, surrogates, and/or proxies Informational: The concept of information as an instrument of national power extends to non-state actors—such as terrorists and transnational criminal groups—that are using information to further their causes and undermine those of the USG and our allies. Military: Fundamentally, the military instrument is coercive in nature, to include the integral aspect of military capability that opposes external coercion. Coercion generates effects through the application of force (to include the threat of force) to compel an adversary or prevent our being compelled. The military has various capabilities that are useful in non-conflict situations (such as in foreign relief). Economic: An economy with free access to global markets and resources is a fundamental engine of the general welfare, the enabler of a strong national defense. In the international arena, the Department of the Treasury works with other USG agencies, the governments of other nations, and the international financial institutions to encourage economic growth, raise standards of living, and predict and prevent, to the extent possible, economic and financial crises.

These instruments of national power are how countries maintain their sovereignty and influence other nations.

In practice these instruments overlap. In particular, informational instruments include public affairs, public diplomacy, communications resources, spokespersons, timing and media. For a long time, the ability to reach mass audiences belonged to the nation-state (e.g. in the USA via broadcast licensing through ABC, CBS and NBC). Now, however, control of informational instruments has been allowed to devolve to large technology companies who have been blissfully complacent and complicit in facilitating access to the public for information operators at a fraction of what it would have cost them by other means.

Democracies and autocracies appear to have different vulnerabilities to information threats [Farrell19][Farrell18][Wooley19]. Democracies require common knowledge (who the rulers are, legitimacy of the rulers, how government works), draw on contested political knowledge to solve problems, and are vulnerable to attacks on common political knowledge. Autocracies actively suppress common political knowledge, benefit from contested political knowledge and are vulnerable to attacks on the monopoly of common political knowledge.

1.2.6 Disinformation for Politics

1.2.7 Disinformation for Power

There are groups who use disinformation for power, but who are not (overtly) part of political parties or geopolitical actions, although they're often directly or indirectly attached to political groups or created/ subverted/ hijacked by geopolitical actors.

Many of these are far-right-wing groups internal to the countries they operate in, but disinformation is always a tempting tool for other activist groups.

There are also groups that use disinformation campaigns to create and use other forms of power. As an example, many of the tactics used by modern power-disinformation groups can be traced back to anti-feminist groups and actions like #gamergate.

1.2.8 Disinformation for Business

1.2.9 Reading

Human vulnerabilities:

- Jonathan Haidt "why it feels like everything is going haywire"
- Demand for Deceit: Why Do People Consume and Share Disinformation? – Power 3.0: Understanding Modern Authoritarian Influence

History of geopolitical influence

- Final Report on the Bulgarian Broadcasting Station New Europe, (Research Unit X.2)
- https://www.psywar.org/articles
- Morale Operations FM
- "Oss morale operations"
- Unrestricted_Warfare
- https://www.psywar.org/content/sibsLecture
- Russian Political War | Moving Beyond the Hybrid

Geopolitical disinformation

- [Farrell19] H. Farrell and B. Schneier “Defending Democratic Mechanisms and Institutions against Information Attacks” Shneier on Security, 2019
- [Farrell18] H. Farrell & B. Schneier “Common-Knowledge Attacks on Democracy” Berkman Klein Center for Internet and Society. Harvard University. October, 2018
- [Wooley19] S.C. Wooley & P.N Howard (eds) Computational Propaganda. Oxford. 2019

Country-specific datasets

- EuVsDisinfo database. Database of pro-Kremlin disinformation https://euvsdisinfo.eu/disinformation-cases/. Ordered by date, narrative, outlets and countries, with summary and disproof. Described in https://euvsdisinfo.eu/old-wine-new-bottles-6500-disinformation-cases-later/. Publicly accessible, no API.
- Facebook GRU dataset provided to SSCI. Not publicly available; described in “Potemkin Pages & Personas”
  Omelas https://www.omelas.io/ has a live feed, multiple countries (Russia, China etc) but I don't think they've gone public with their dashboard yet - can ask for email summaries
- Russia analysis: KremlinWatch does analysis on Russia-EU ops https://www.kremlinwatch.eu/#welcome ; CEPA is more high-level http://infowar.cepa.org/This-week-in-infowar.
  If you’re looking for non-Russia, you’re basically looking at specialists.

Disinformation from the Defender POV

What disinformation targets

Disinformation uses people the way that malware uses PCs. Sometimes people, and clusters of people (communities, nations etc) are the endpoints, and sometimes they're channels (e.g. influencers, media) to reach more people, to spread narratives, create confusion or increase community fragmentation and distrust.

Countries sometimes target other countries to weaken them by helping populations distrust each other and their systems and officers of governance, and act in ways counter to a strong nationstate. Countries also target their own populations, e.g. attacking the credibility of non-ruling parties, voting systems or minorities to stay in power. Successful gambits

include increasing distrust between internal groups, often by targetting disinformation campaigns at one or all of the groups around a divisive debate.

Fraudsters target anyone who will give them money. Often this is as simple as building campaigns around getting eyeballs onto a sales site (or just a website: eyeballs and clicks are worth advertising money), by piggybacking on divisive or emotionally-charged conspiracy narratives like Covid5G.

There has been some directly targetted disinformation. Individuals (BillGates, Fauci) have had targetted disinformation campaigns around them; some campaigns directly targetted hospitals as part of the "covid isn't real" narrative, and some companies have used disinformation to alter rivals' prospects. Some hybrid infosec/disinformation attacks using deep faked voice also exist but are still relatively rare compared to e.g. ransomware. Commercial disinformation appears at the moment to be generally spam and marketing companies pivoting to disinformation as a service as a new line of business.

Big, Fast, Wierd: why disinformation is getting harder to track

When we talk about security going back to thinking about the combination of physical, cyber and cognitive, people sometimes ask why now? Why, apart from the obvious weekly flurries of misinformation incidents, are we talking about cognitive security now?

One answer is the three Vs of big data: volume, velocity, variety (the fourth V, veracity, is kinda the point of disinformation, so we're leaving it out of this discussion).

- Variety: The internet has a lot of text data floating around it, but its variety isn't just in all the different platforms and data formats needed to scrape or inject into it — it's also in the types of information being carried. We're way past the Internet 1.0 days of someone posting the sports scores online and a bunch of hackers lurking on bulletin boards: now everyone and their grandmother is here, and the (sniffable, actionable and adjustable) data flows include emotions, relationships, group sentiment (anyone thinking about market sentiment should be at least a little worried by now) and group cohesion markers.
- Volume: There's a lot of it — volumes are high enough that brands and data scientists can spend their days doing social media analysis, looking at cliques, message spread, adaption and reach.

- Velocity: And it's coming in fast: so fast that an incident manager can do AB-testing on humans in real time, adapting messages and other parts of each incident to fit the environment and head towards incident goals faster, more efficiently etc. Ideally that adaptation is much faster than any response, which fits the classic definition of "getting inside the other guy's OODA loop".

NB The internet isn't the only system carrying these things: we still have traditional media like radio, television and newspapers, but they're each increasingly part of these larger connected systems.

Another common question is "so what happens next". One answer is to point people at two books: The Cuckoo's Egg and Walking Wounded — both excellent books about the evolution of the cybersecurity industry (and not just because great friends feature in them), and say we're at the start of The Cuckoo's Egg, where Stoll starts noticing there's a problem in the systems and tracking the hackers through them.

We're getting a bit further through that book now. In America, if someone sees a threat, someone else makes a market out of it. Cuddle-an-alligator — tick. Scorpion lollipops in the supermarket — yep. Disinformation as a service / disinformation response as a service — also in the works, as predicted for a few years now.

Disinformation response is also a market, but it's one with several layers to it, just as the existing cybersecurity market has specialists and sizes and layers. One of the reasons for working on disinformation threat intelligence is to help encourage that market to grow.

Readings

- The Cuckoo's Egg
- Walking Wounded
- Rent-a-troll: Researchers pit disinformation farmers against each other
- Market Sentiment

chapter-3_aa.md

description: CogSec: Deconstructing Disinformation

Chapter 3_aa

Cognitive Security

Information Security has always had three main layers (cognitive security, physical security, cybersecurity), but the cognitive one has been downplayed for a long time. Cognitive security is a rapidly growing domain that interacts with cyber and physical security, and includes things like information operations and disinformation. This covers tools, techniques and resources for threat sharing and response, and practical applications.

Ah. Yes. So there's a name clash here - some groups use "cognitive security" to mean using AI in information security defence (FWIW, it's not often heard in MLsec communities). But here, we're talking about information security as having three main layers: cybersecurity (the machines and networks infosec you usually read about), physical security (physical, as in breaking into the building because it's sometimes easier to steal from the computer than break in electronically) and cognitive security (attacking and defending human minds and the networks between them as part of an infosec attack).

Disinformation is an attack in the cognitive security domain, but there are others that can be used - social engineering is getting humans who are part of an information security system to help you break that security (e.g. by letting you into their systems). The work that we at misinfosec and others did on how attacking human beliefs and emotions can be viewed and defended in similar ways to attacks on machines is, broadly speaking, the Cognitive Security field (this one, not the other one...).

Disinformation Layers

Disinformation pyramid

As we explore and analyze the information sphere, analysts have techniques that are employed to understand disinformation operations - and they're classified using similar frameworks to those we use to classify our understanding of other types of threats and incidents.

- Campaigns: are long-term disinformation operations. They're focussed around a theme, like specific geopolitics (e.g. "make everyone like china" or "Ukraine is really Russia"), and are often nation-state-funded, but might also be from interest groups (e.g. far-right-wing, antivaxxers etc).
- Incidents: these are the short term, cyclic things we track. They're coordinated sets of activities that happen over a defined timespan that usually indicates some form of team or individuals driving them. Incidents have things with defined parameters like TTPs that we can share, threat actors, and other objects that you'd recognise from TI, but also including context and narratives.
- Narratives: are the stories that we tell about ourselves and the world. They're stories about who we are, who we do and don't belong to, what's happening, what's true (e.g. Covid19 was caused by 5G masts). Tagging information with defined narratives make

it easier for us as analysts to follow the flow of information across the internet and beyond.

- Artefacts: Incidents and Narratives show up online as artefacts: the text, images, videos, user accounts, groups, websites etc and links between them all that we collect and use to understand what's happening.

So what looks to outside observers like analysts simply hunting down a hashtag or a URL, describing a narrative, or trying to understand the things that link to it is so much more; it's really a part of creating an inventory of the discrete elements of each incident, or the objects used by a disinformation team or campaign, so we can a) share a summary of what we think is happening, and b) disrupt both those component parts, the TTPs behind them, and the incidents and campaigns they support.

This is a lot of text. And we're realising that there's a lot of stuff we haven't explained. So we're writing it down. And making stuff clearer and cleaner to use as we test and explain it. This document is those explanations.

Disinformation Objects

STIX and extensions...

Disinformation Tactics, Techniques, Procedures (TTPs)

AMITT Framework

misinformation-tactics | Analysis | Initial

Strategic Planning (4 items)	Objective Planning (2 items)	Develop People (3 items)	Develop Networks (6 items)	Microtargeting (3 items)	Develop Content (10 items)	Channel Selection (10 items)	Pump Priming (8 items)	Exposure (10 items)	Go Physical (2 items)	Persistence (3 items)	Measure Effectiveness
5Ds (dismiss, distort, distract, dismay, divide)	Center of Gravity Analysis	Create fake Social Media Profiles / Pages / Groups	Create hashtag	Clickbait	Conspiracy narratives	Twitter	Bait legitimate influencers	Use hashtag	Organise remote rallies and events	Continue to amplify	
Competing Narratives	Create Master Narratives	Create fake experts	Cultivate useful idiots	Paid targeted ads	Adapt existing narratives	Backstop personas	Demand unsurmountable proof	Cheerleading domestic social media ops	Sell merchandising	Legacy web content	
Facilitate State Propaganda		Create fake or imposter news sites	Create fake websites	Promote online funding	Create competing narratives	Facebook	Deny involvement	Cow online opinion leaders		Play the long game	
Leverage Existing Narratives			Create funding campaigns		Create fake research	Instagram	Kernel of Truth	Dedicated channels disseminate information pollution			
			Hijack legitimate account		Create fake videos and images	LinkedIn	Search Engine Optimization	Fabricate social media comment			
			Use concealment		Distort facts	Manipulate online polls	Seed distortions	Flooding			
					Generate information pollution	Pinterest	Use SMS/ WhatsApp/ Chat apps	Muzzle social media as a political force			
					Leak altered documents	Reddit	Use fake experts	Tertiary sites amplify news			
					Memes	WhatsApp		Twitter bots amplify			
					Trial content	YouTube		Twitter trolls amplify and manipulate			

Select Some Options

AMITT, as seen in MISP

We're using the AMITT framework to break each disinformation incident down into its component TTPs and TTP-level counters. AMITT (Adversarial Misinformation and Influence Tactics and Techniques) is a framework designed to give responders better ways to rapidly describe, understand, communicate, and counter misinformation-based incidents.

AMITT is designed as far as possible to fit existing infosec practices and tools, giving responders the ability to transfer other information security principles to the misinformation sphere, and to plan defenses and countermoves. The latest version of AMITT is held in the Github repository https://github.com/misinfosecproject/amitt_framework

The language and style of the AMITT framework is adopted from the MITRE ATT&CK framework. The framework is read left-to-right in time, with the entities to the left typically (but not necessarily) happening earlier in an incident. The phases are separated into left-of-boom (purple) and right-of-boom (red), to represent activities before (left) and after (right) an incident is visible to the general public. Every AMITT component has a unique id (e.g. T0018 Paid targeted ads). To use AMITT, list and share the components you see in your incident, e.g.: Phases (top row: purple and red boxes): higher-level groupings of tactics, created so we could check we didn't miss anything. The tactics below each phase belong to that phase. Tactics (second row: blue boxes): stages that someone running a misinformation incident are likely to use Techniques (all other rows: grey boxes): activities that an incident creator might use at each stage. The techniques below each tactic belong to that tactic. Tasks (not shown): things that need to be done at each stage. Tasks are things you do, techniques are how you do them. Compiling and reporting incidents is an important aspect of both responding and developing the tools needed to do so. To be effective, those reports should include as much information as possible about the stages and techniques at play in those incidents. We've created an html tool to help you with this https://github.com/misinfosecproject/amitt_framework/blob/master/matrix_to_message.html - download it, and click on the boxes that you need in your report. You can then cut-n-paste the list of techniques that it generates at the bottom of the page.

Reading

AMITT

- Walker et al, Misinfosec: applying information security paradigms to misinformation campaigns, WWW'19 workshop

Reading

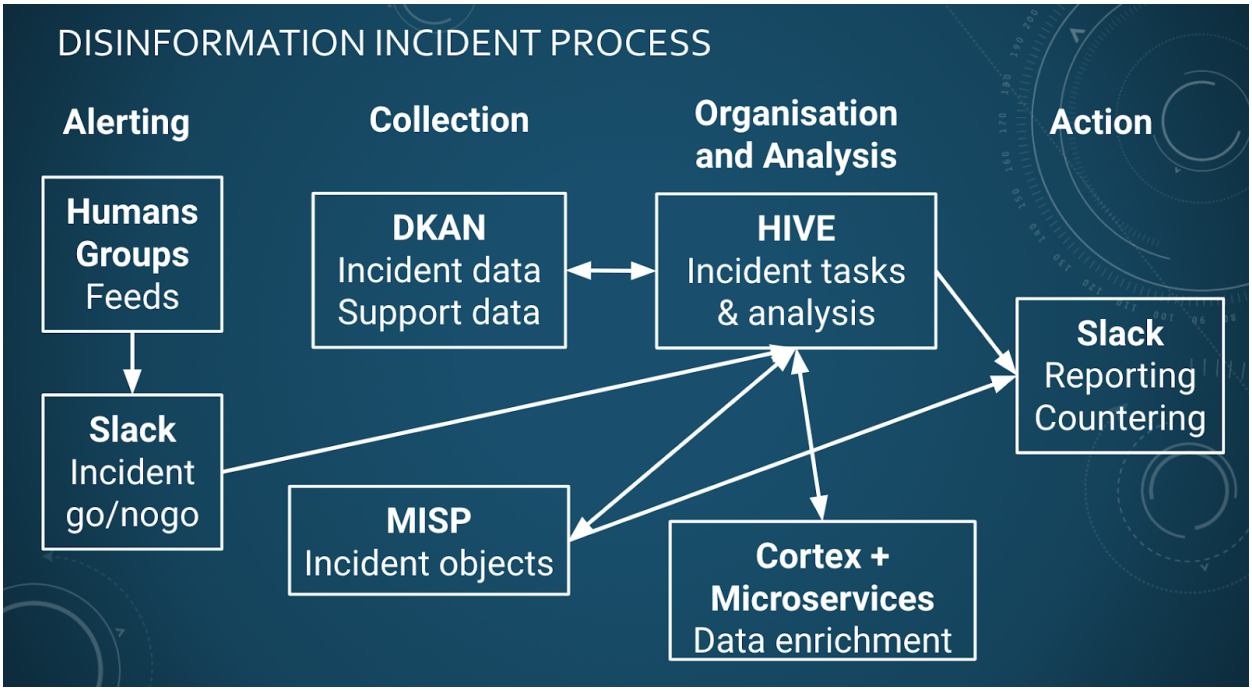
- Russian Election Trolling Becoming Subtler, Tougher To Detect
- Big Lies and Rotten Herrings: 17 Kremlin Disinformation Techniques You Need to Know Now
- http://overcognition.com/2019/05/13/misinformation-has-stages/
- https://medium.com/misinfosec/disinformation-as-a-security-problem-why-now-and-how-might-it-play-out-3f44ea6cda95

Chapter 4

Disinformation workflows include:

- Tracking an incident
- Tracking narrative flows across incidents
- Adding and maintaining supporting disinformation data

1.1 Incident workflow



Incident dataflow

The main workflow in the disinformation team is tracking an incident. We've broken this into 4 stages: alerting, collection, analysis, and action.

1.2 Workflow instructions

Starting a disinformation incident:

A new Covid19-related rumour has started online. You've seen it yourself, someone has sent you an example of it, you've seen another group tracking it - there are a bunch of ways to spot something new happening. Now what? NB each of these steps can be by different people

1. Tell people
    1. Put a message in slack #4-disinformation, with the artefact you found and a short description.
        - Start with "NEW RUMOR" so we will be able to track them
        - Any supporting information or links (under that rumor) should be posted in a thread off that initial NEW RUMOR post
        - This will make documenting and adding objects and observables to the incident and analysis log easier to track, and also keep everything a little more tidy
2. Decide whether to start an incident
    1. Do a quick check that it's a rumour. One sighting doesn't make an incident. 15 copies of the same message on Twitter, or 3 friends sending you the same strange DM, and you're probably onto something.
3. If it's significant, start an incident
    1. Give it a name. Names help.
    2. Add a row to the [incidents spreadsheet]
    3. Create a folder in the [googledrive INCIDENTS folder] for notes and anything that won't fit into the DKAN
    4. Start adding data to the DKAN ([learn more about DKAN in the Tools chapter]
4. Investigate the rumor
    1. Look for related artefacts, accounts, urls, narratives etc
5. Investigate ways to close down the rumor / repeater sites etc.
6. Report on the rumor
    1. Add an incident to the MISP instance for this rumor ([learn more about MISP in the Tools chapter]
        - The incident must include some relevant observables such as a Tweet, social media username or URL.
    2. Write and send notes/reports to the people who can respond
    3. Close down the rumor and move onto the next one (there's always a next one)

Help with a disinformation incident

1. The master document for what we're doing on incidents is the [incidents spreadsheet]. Look at the status column - the priority is live incidents, then monitor long-term, then "keep an eye on it" (the potential 'zombie' incidents that are probably dead but might restart)
2. Check back in the slack channel, and in the incident README in the [googledrive INCIDENTS folder] to see what's been done with this incident recently. As we get things together, we'll probably have incident-specific tasks in the github issues list, but we're still working on that.
3. Find articles and artifacts, investigate the ones we have, put results into the slack channel for harvesting by the bots, and/or discussion with the team.
4. If you spot something significant (new objects tied to the incident etc, new things of interest), update the incident README.

1.3 Alerting

A team has many places it can potentially get disinformation alerts from. These include:

- Alerts from disinformation team members
- Feeds from other groups
- Phone honeypots
- Reporting hotline (dedicated email address)
- Sniff disinformation report lists, dashboards and botnet feeds for themes
- Set up reporting from social media (Facebook, twitter etc)
- Ask social media companies for feeds from them
- New data coming into the DKAN

We learn about potential incidents from several places:

- Teams connected to this one, e.g. Covid19activation and covid19disinformation, who are watching for disinformation online
- Team members spotting online disinformation and raising the alert in the team slack channel
- Team members spotting alerts from other disinformation tracking teams

- Other CTI channels telling us about disinformation in their feeds

Important: An alert isn't the same as an incident. An incident needs to be within the team's scope, and large enough to be worth expending team effort on.

When we see an alert, we have some questions:

- Is this an incident, e.g. is it a large coordinated disinformation incident, or an isolated piece / few pieces of disinformation?
- Is this disinformation suitable for processing by the disinformation team (e.g. 419 scams might be better handled by the Phishing team, but might also contain information about incidents that we should check out too)?
- Is this disinformation already being handled by platform teams or other specialist teams (we might want to check in with them just in case, for instance referring to healthcare groups ior law enforcement, or issuing a takedown request because of a finding)?
- Is this incident something that we should track?

"Is this incident something that we should track?", e.g. how do we choose which incidents to track?

- We don't track incidents for fun or interest. We track the ones that we have a reasonable chance of doing something useful about - whether that's raising the alarm to groups or organisations that can respond to the incident, asking them to take specific actions (like taking down a disinformation account or site), or taking actions ourselves (like amplifying counternarratives).
- We also track and counter incidents that we believe give us the best chance of a positive effect, and in the Covid19 deployment, ideally one that impacts health.
- Yes health. We prioritise that over other incidents, although we will include disinformation around current events where they impact populations.

1.4 Hive lists for starting an incident

When you create a disinformation incident in HIVE:

- Create a new case. Use case template “Influence Operation Incident”.
- Name the incident (use this name in all the tools)
- Create an event in MISP for the incident
- List the risks and potential real-world consequences from this incident
- List any time bounds on the incident, e.g. are there events that it’s gearing towards etc
- List any geographical or demographic targets in this incident
- Create a DKAN directory for the incident

MISP list for starting an incident

- List actors and other objects that are important in this incident - we’re using a combination of STIX and DFRlab’s Disinformation Dichotomies standard for this. Add these to the Clean MISP
- List the tactics and techniques that are being used in the incident - we’re using AMITT for this (the version that comes as standard in MISP). Add these to the MISP event.

1.5 Organisation

Documenting analysis:

- We have DKAN and MISP, but also useful to have a google folder for each incident for other things that don’t fit into those, like research notes
- Classifications: if it’s openly available online, then it’s okay to put through e.g. Tableau; if it’s come though internal routes (e.g. SMS), then keep it off public internet (don’t share).
- looking for related artifacts, urls, narratives etc

Who we communicate to:

- Report when something significant happens - e.g. see this main effort for this new line
- Report on time period… if big, a daily report; if smaller a weekly report
- No report goes out without at least 2 people beyond the editor going over it
- End users are also watching the MISP

Who makes decisions:

- Depends on decisions
- Need a board - vote via slack; person calling for vote does @channel to board, or emails them
- Who can add an incident? Anyone can start an incident.
- Who can release a report -
- Who can talk to customer/ victim? Needs to be agreed on

1.6 Collection

DKAN holds data we don't want to lose, and data that's raw and large: it's the in-tray

MISP hold objects of interest and the relationships between them, so we can quickly look up things we've seen before etc

Data we build up in MISP

- Incidents
- Narratives
- Actors
- Urls

1.7 Action

What we want to do with an incident is disrupt it as much as possible. If we can stop it completely, that's a big win, but generally, we're after disruption. CogSecCollab has a long-list (here: https://github.com/cogsec-collaborative/amitt_counters/blob/master/tactic_counts.md) of the things we can do to disrupt incidents at different stages of the disinformation killchain (https://github.com/cogsec-collaborative/amitt_framework - that, and DFRlab's object labels https://github.com/DFRLab/Dichotomies-of-Disinformation are what we're using in the MISP reporting), but frankly it's still messy so at this stage it's better to put our hacker hats on and think "which artefacts (observable objects) do we have in this incident, and what can we do to make them less effective?"

Examples: are there URLs pushing out covid5g disinfo? Are there social media accounts and groups pushing out covid5g disinfo? If we gather evidence on these, we can get that to the social media companies. Are there botnets involved (yes, yes, I said the b word, but they're part of this too)? Can report those too. Etc etc (and I suspect many of you have etcs CogSecCollab didn't think of when they created that counters repo).

This is the practical part of incident handling. We track an incident until the underlying incident stops or slows significantly (or the event it's building up to has passed), or until we've done as much as we believe we can to counter it, or know that there are other teams dealing with it.

Disinformation counters are much more than "remove the botnets" and "educate people". For most incidents, there are a variety of things that can be done about the incident, its creators, the objects used in it, and the tactics and techniques used. We've collected a few (well, a couple of hundred) suggestions for technique-level counters at https://github.com/cogsec-collaborative/amitt_counters - we're expecting to uncover a bunch more as more infosec people do disinformation.

1.8 Managing an incident response

An individual can track an incident on their own - open up some notebooks, fire up the coffeemakers and mainline chocolate for a couple of days. That's - not sustainable over time and large numbers of incidents, any more than it is for other infosec incidents.

The short instructions for managing a response are in the [team readme]. This is some of the thinking around them:

We haven't worked out exactly how to fit cognitive security / disinformation response into a SOC yet, but here's where we are at the moment on starting an incident:

- Incidents need names. Yes, yes, I know that's a slippery slope that ends up in a cute mascot and a dedicated website, but a name makes it easy to quickly identify what you're working on, find the right folder to put things into etc.
    - Action: Make up a name: make it short but descriptive - you're going to be typing it a lot, but you also want to remember what it was about a week later.

- The team needs to know you started an incident - both the team who are around at the time (and can help look for artifacts, add their specialist skills etc), team members who are coming in looking for things to do later, and leads who are trying to balance the load on the team overall. Best way to do this is to add a note to the team chat and an entry in the team log.
    - Action: add a note to the team slack channel, naming the incident and asking for help with it (if needed). If you have a starting artefact, add that too. Adding the word “NEW” will make it easier to find by people looking in on the channel later.
    - Action: add an entry in the team log, saying you’re starting an incident response. At the moment, this is the incidents spreadsheet - this is likely to shift to adding a case to an incident tracking tool like TheHive.
- You, and the team, are going to start producing notes and artifacts as you track through the incident. Create a place to put them, that’s accessible to the team
    - Action: create a space to put images, artifacts etc in. At the moment, that’s creating a folder for the incident under the INCIDENTS googlefolder - this is likely to shift to directly uploading to a tool like TheHive or MISP.
    - Action: create a notes log for the incident. At the moment, that’s a README file in the incident googlefolder - this is likely to stay the same for the moment. In the log, write a short description of the incident, and how you started tracking it (e.g. what the first artefact(s) you saw were.

Here’s where we are on managing investigating the incident:

- You, and the team, are going to investigate the incident
    - Action: Look for related artefacts, accounts, urls, narratives etc
    - Action: add artefacts to the space you set up for collecting images, artefacts etc. You’ll find it helpful if you number the images, because they’re difficult to reference otherwise (aka “the yellow poster again” isn’t as specific as “image001_yellowposter”)
    - Action: keep the flow of investigation moving - keep a list of actions related to the artefacts, and/or direct the team to areas that need further research
- You’ll also need to translate that into an incident description that can go out as an alert to other teams, and be used to look for potential counters
    - Action: add incident to alert tools. We’re using MISP here, so adding a MISP object for the incident, and attaching the objects important to it is appropriate here.
    - Action: map artefacts seen to tactics and techniques. MISP includes AMITT - you can use the ATT&CK navigator to click on all the tactics and techniques you can

see in this incident.
    - Action: Investigate ways to close down the rumor / repeater sites etc. We're working on tools for this too, but for now it's discuss this with the team, and check the lists below.
- Oh, and yes, you get to be scribe for the team too, making sure you keep a record of the investigation:
    - Action: keep the incident log updated with any significant findings, notes, things to do etc.

And here's where we are on managing responding to the incident:

- You need to get information about the incident out to other teams that could do something about it:
    - You've already added an incident to MISP; make sure it's ready to go (question: is there something we need to do to get it out on the feeds?).
    - Write and send notes/reports to the people who can respond
- If you found ways to respond, decide what to do, and check whether you did it
    - If the team found ways it could respond - triage them. Find ways to do the ones you can.
    - Also check on the things you were going to do. Was something done? Chase it up.
- And finally, know when to stop.
    - If you've done as much as you sensibly can, close down the rumor and move onto the next one (there's always a next one).

There are always more incidents, although we're often lucky enough to have a few days without anything major going on. Every morning, one of the leads looks through the list of incidents and decides which ones should continue to be 'live', which we should move to just keeping an eye on, or keep a longer-term watch on in case they flare up again, and which we can close down as unlikely to be active again.

Chapter 5

Narrative workflows

Narrative: Narratives are part of incidents - each incident might have multiple narratives involved, or just one, but there's usually an identifiable narrative somewhere in there, that you can use to see if there are related incidents already tracked or dealt with etc.

But there are a lot of them. Hence the mindmap, which starts to group narratives into hierarchies, making them easier to read and manage.

The other thing about narratives is that they, like incidents, have lifetimes. Some narratives appear as a result of a world or local event (or upcoming or anticipated event), and are only useful whilst that event is in peoples' minds. Example: using the Stafford Act to make everyone stay indoors was a narrative we tracked a month ago, before the stay-at-home orders started and it was a lot clearer about what states could, couldn't, would and wouldn't do.

Other narratives appear for a while, go dormant, then reemerge in different forms. Example: 5G, which was originally part of the radiation-of-all-forms-will-do-bad-things-to-you narratives, and has now come back in a mixup with covid19.

So what we need is a way to log all the narratives that we know (or care) about, whilst keeping a smaller list handy of "currently alive" narratives that we can check incoming disinformation against.

Identifying new narratives

Part of our work is to identify new threats before they become widespread. One way to do this is to identify emerging narratives from our existing asset collection.

First, we need to establish a baseline understanding of the current threat landscape in our area of interest (e.g. anti-mask, covid5g etc). The places we look to start this work include:

- Master narratives lists
- Existing lists of persistent threats known to carry disinformation: known bots, sources (e.g. disinformation websites), and canaries (accounts or hashtags with a high probability of carrying disinformation in this area)
- Regular threat streams: known disinformation feeds, subscriptions and platforms.

Once we have a baseline, we can establish persistent and repeatable monitoring:

- Identify data sources to monitor, e.g. googlenews, twitter, facebook, news aggregation sites etc
- Create saved or formatted searches for each platform, e.g. twitter = '#disinformation covid qanon boogaloo'; google = google hack formatted with a time parameter, e.g. 'disinformation and covid when=1d'
- Where api access is difficult, use other platform collection resources where possible, e.g. tweetdeck, crowdtangle

Other ways to find outlier or new narratives include watching for one or more of:

- merging and/or reemerging narratives being pushed by usually opposing groups, or old narratives that are reactivating
- local or world events, e.g. protests, changes in an area's status around specific dates (holidays etc)
- anomalous or significantly-sized online activity, e.g. in trending hashtags

Once narratives are found, you'll need to analyse them:

- evaluate source biases (is this state-owned media, an opinion article, social media etc)
- find additional sources with the same and/or competing narratives
- compare and contrast your findings: what's the same - is this fact or opinions? What's different - why? What's the intent and/or agenda behind the narrative - is it political, influence, harm, designed to confuse, distract, disrupt?
- How could this be used for bad (you might want to red team this)
- What would the impact be if this narrative is leveraged for bad?

Chapter 6

1.1 Data inputs: Alerts and Canaries

We receive alerts about possible disinformation incidents from members of the disinformation team, and from other teams connected to us. Typically we get alerts around an artefact or theme, e.g.

- A new narrative emerging online, either in general social media or known conspiracy / extremist / target etc groups
- A local or world event that might spark a disinformation incident
- Anomalous or significant-sized online activity that might be associated with a disinformation incident
- Command signals from known disinformation groups (e.g. qanon)

The types of artefact that we typically receive include:

- Images
- Messages, e.g. tweets, facebook posts, SMS or Messenger/Telegram etc messages
- URLs

The processes for investigating these are discussed in more depth in the next chapter.

Several accounts and groups are either known producers or early adopters of many disinformation campaigns. We've dubbed these "canaries", as in the entities that give the first signals that something is happening (canary, as in "canary in a coal mine").

1.2 Data sources: disinformation data streams

When we get our first data inputs, it's a good idea to check them against other disinformation and related data collections, to see if they've been picked up by other researchers, or those researchers have already collected data related to these inputs that can be of use to our investigation. The data feeds are continually updated, so are a good

source for breaking data; the static data collections are good for finding history on data, source, narratives etc.

1.2.1 covid19-related disinformation data feeds

Narratives

- Wikipedia list of Covid19 rumours
- WHO Covid19 myths list - narratives
- EuVsDisinfo database https://euvsdisinfo.eu/disinformation-cases/
- Ryerson Claimwatch dashboard
- CMU IDEAS Center list of Covid19 disinformation narratives (click dates)
- Indiana Hoaxy (twitter, articles)

Data

- Botsentinel: lists “trollbots” (bot-like and troll-like accounts) and the themes they’re promoting https://botsentinel.com/ (not just Covid19)
- Hamilton68 - live feed from accounts attributable to Russia or China (may or might not contain propaganda; useful for seeing current themes). Public version is live feeds from official Russian sites (embassies, RT etc), not trolls. Academics can ask for a more detailed feed. https://securingdemocracy.gmfus.org/hamilton-dashboard/ (not just Covid19)
- Ryerson University covid19 misinformation portal: https://covid19misinfo.org/
  - Botswatch dashboard https://covid19misinfo.org/botswatch/
- Uni Arkansas COSMOS Covid19 list http://cosmos.ualr.edu/misinformation
- Indiana University OSOME Decahose
- Facebook datafeed: Enabling study of the public conversation in a time of crisis

Domains

- Coronavirus Misinformation Tracking Center – NewsGuard

1.2.2 Covid19-related counter-disinformation feeds

- Ryerson University covid19 misinformation portal: https://covid19misinfo.org/
- Snopes: https://www.snopes.com/

- WHO COVID-19 site: https://www.who.int/health-topics/coronavirus
- WHO information network for epidemics https://www.who.int/teams/risk-communication
- Coronavirus Tech Handbook https://coronavirustechhandbook.com/misinformation
- Experts list https://twitter.com/jeffjarvis/status/1254038157244456961
- Maryland Covid19 rumour control https://govstatus.egov.com/md-coronavirus-rumor-control

1.2.3 Covid19 general data feeds

- https://crisisnlp.qcri.org/covid19 - GeoCov19 dataset of covid19 tweets (up to about 3 weeks ago; still collecting)

1.2.4 General disinformation datasets

- Twitter IO archive: covers several countries up to a few months ago. Good for getting a sense of the size and 'feel' of typical nationstate twtter posts/ networks etc. https://transparency.twitter.com/en/information-operations.html
- Facebook ad library: contains all active ads that a page is running on Facebook products https://www.facebook.com/ads/library/ (About the Ad Library)

1.3 Collecting your own data using tools

The datastreams above will help you get a sense of what's known about the artefact and/or theme that you're investigating, and sometimes that's enough to craft a response (e.g. if there's a WHO page on a known scam, that might be enough evidence to ask for takedowns etc). But most of the time, you'll have to go collect your own data from across social media, and sometimes beyond (e.g. for paper flyers, we asked people if they'd seen them in their neighbourhoods too).

Where you collect from, and what you collect will depend some on the artefacts you found, but here are some of the ways.

1.3.1 Twitter data

Twitter data is studied a *lot* precisely because it has a lovely API. Since we use a lot of Python here, let's talk about Python libraries. If you have twitter API codes, then Tweepy is a good choice. If you don't want to use the twitter API, try Twint.

Various researchers post twitter data-gathering tools online. Andy Patel's twitter-gather is good if you're doing twitter network analysis. We have code based on an early version of this in the github repo. It's andy_patel.py - call it with "python andy_patel.py name1 name2 name3 etc" where name1 etc are the hashtags, usernames, phrases (phrases in quotes) that you want to search Twitter for. Andypatel.py creates a set of files in directory data/twitter/yyyymmddhhmmss_hashtag1 etc with the tweets, most prolific urls, authors, influencers, mentions etc and gephi input data so you can create user-user etc graphs (see the gephi instructions in this BigBook for how to do that). Data for earlier investigations are in the repo folder data/twitter if you want to see what that looks like.

1.3.2 Facebook data

The Facebook API is horrible. Most everyone tracking social media uses a third party like CrowdTangle (which isn't free) or scrapes for the data they want.

1.3.3 Reddit

Reddit data is regularly dumped in an easy to read format. For quick-looks, there are tools like https://www.redective.com/

1.3.4 Multi-platform tools

Reaper collects from a set of social media feeds. Trying that out.

Access tokens:

- Facebook: look at list in https://developers.facebook.com/docs/facebook-login/access-tokens/ - then used https://developers.facebook.com/tools/explorer/ to check token worked before putting into reaper.
    - "Page Public Metadata Access requires either app secret proof or an app token" - see https://developers.facebook.com/docs/apps/review/feature#reference-PAGES_ACCESS

Storing datasets

Social media data can be large, and its value is often in the relationships between objects as well as the objects themselves. Options we've used include collections of CSV and json files held in a DKAN data warehouse, Neo4j (https://neo4j.com/download-neo4j-now/) and an ELK stack (https://www.elastic.co/)..

Chapter 7

Artefacts are the things we can see online - they're what we track and use to understand what's happening in an incident, how everything in it fits together, and what we can usefully pass on as information about it at the incident level, or usefully do to influence it. The artefacts that we see most often include:

- Tweets
- Twitter accounts
- Facebook groups
- Domains - websites
- Hashtags
- Images
- Videos
- Audio fragments (e.g. voice messages)
- yas

The next layer up includes:

- Narratives
- Botnets

The basic questions: What is this thing. How is it impacting the things we care about? Are there other teams doing something about it? What can we do about it? How much impact can we make in the things we care about, for the resources we need to expend?

1.1 Handling Domains (URLs)

1.1.1 Chasing a URL

So you've got a URL. Now what? Well, you probably want to know about the URL - who created it, when, what's it connected to etc.

Check for company

- Is anyone else tracking this url? Check reddit etc. - you might save yourself time if other groups have already tracked lists, social media etc.

All the Google Dorks! (h/t to Roger)

Using Google Dorks to Check Primary sources (from Henk van Ess's Finding patient zero)

Websites as primary sources: This is useful when your searches within specific sites or urls are coming up empty

Step 1: Look at the failing link

- Ex. https://www.sec.gov/litigation/apdocuments/3-17405-event-11.pdf
- Pull out just the domain name and Top Level Domain (Ex. sec.gov)

Step 2: Use "site:"

- Go to a generic search engine.
- Start with the query ("Dutch police") and end with "site:" followed directly with the URL (no spaces).
- Ex. "Dutch police" site:sec.gov

Step 3: Adapt the "primary source formula" to your needs

- Include specific folders (Ex. "Dutch police" site:sec.gov/public)
- Predict folders you think might be there

Following the trail of Documents

Step 1: Establish the document type

- Is it a doc | pdf | xls | txt | ps | rtf | odt | sxw | psw | ppt | pps | xml file?
- Use filetype: and the type of file with no spaces (Ex. "filetype:pdf")

Step 2: Include a phrase you'd like to search with in the document (could include a date)

- Ex. You're searching for an invitation to an event from May 13, 2014, event. (Be sure to search for both the cardinal and ordinal forms, May 13 and May 13th.)

Step 3: Who is involved?

- Do you know the creator/host and it's website?
- Ex. The organizer is "Friends of Science" and its website is friendsofscience.org.

When you combine all three steps, the query in Google will be:

"May 13th, 2014" filetype:pdf site:friendsofscience.org

Filtering social media for primary sources

YouTube

YouTube's search tool has a problem: it won't let you filter for videos that are older than one year. To solve this,

- In a Google search include the keywords and site:youtube.com
- manually enter the preferred date into a Google.com search by using the "Tools" menu on the far right
- Then select "Any time" and "Custom Range."

Process for investigating the authenticity of a website :

Web searching a domain: Since we want to find out what other sites are saying about the site while excluding what the site says about itself, we use a special search syntax that excludes pages from the target site

- Search syntax is website -site:website
- (Ex. baltimoregazette.com -site:baltimoregazette.com)
- Scan the set of results looking for sites we trust

Finding out who runs a site with WHOIS :

- Enter the domain name into the WHOIS Domain Tools or https://lookup.icann.org/lookup
- Note who the domain was registered to
    - Unfortunately, WHOIS blockers have dramatically reduced the value of WHOIS searches, so you may only find a proxy.

- Note when the domain was registered

Use a backlink checker like ahrefs or smallseotools that allows you to see all websites that link to a particular site

Look up the url

- Builtwith:
  - look on builtwith.com - if you're lucky that will tell you when and who
  - It will also tell you which sites have the same tags as this site: this helps you find connected sites
  - Use CSC code run_builtwith.ipynb - same thing, but gives you json and a dataframe of those connected sites

Look at the URL contents

- Are there phrases you can use in a googlesearch, to find related objects? Run the search that allows repeated results, to see identical pages. About and terms pages are usually good places to look for these.
- Use CSC code googlesearch_for_terms.ipynb to search for terms/ pages.
- Are there people connected to the site? Start searching for them
- Are there companies?

Think about geography etc

- Look at the title and url of the site. Do they have elements that might be repeated? E.g. if you have xxxmichigan.com, check for the same pattern with other states' names, e.g. xxxwisconsin.com. Astroturfers try to cover an area, whether it's geographical or demographic, and if they're doing it for money, they'll usually have multiple sites.

Look for links

- Check social media - are there references to the URL, or groups / pages / accounts with the same name?
- If there are references to the URL, are there common hashtags, phrases or people in common you can use to search for more sites?

Examples:

- "Data Safari rough notes: “pink slime” network"

1.1.2 Look for ‘similar’ websites

Typosquatting is when you create a site whose url is *almost* the same as a real or well-known one, often using combinations of letters (e.g. ‘nn’ instead of ‘m’) or urls (e.g. .gov.us) to fool people on a casual glance.

Useful python libraries for generation typosquats include dnstwist

- Near-duplicates SnaPy
- typosquatting

Feed of domains that were created each day: whois newly-registered-domains

- Idea: we could search this feed each day for domain names matching the things of interest to us, e.g. MMS

1.1.3 Look for new sites

Github code check_new_registrations.ipynb searches for strings of interest in newly-registered domains (from whois newly-registered-domains). But newly registered alone isn't really an indication of anything; domains that are newly registered and active all within 24hrs, are worth watching, as is any recently active and questionable domain. We have e.g. the Zetalytics API for searching through those.

1.1.4 Social media references to the site

Crowdtangle chrome extension will give you a list of references to a site you’re looking at, on Facebook, Twitter, Instagram and Reddit.

1.2 Handling Tweets

1.2.1 Chasing a hashtag

“what do we consider worthy of collecting from twitter?” - FrankC

Good question. The TL;DR is that the reason we use the code that we do (andypatel_gettwitter.py from CSC tracking repo) is because we’re looking for the objects that dominate and are related to the hashtag:

- we want to know which users are promoting it
- Which other hashtags are used heavily with it
- Which users on the hashtag are in suspicious configurations - e.g. one user linked out to lots of other people who aren’t connected to each other (that’s someone either pushing or pulling, depending on the direction of the links), or groups of users connected heavily to each other but not to anyone else on that hashtag (typical configuration for a botnet)
- we want to know which URLs are associated with the hashtag - if this is being used to make money, that money has to come from somewhere, and that’s usually either online advertising, merchandise or paid services: either way, each of those is going to have a web address associated with it, and any grifter worth their salt is going to be pushing that address heavily
- We also collect images - that gives a good idea of what the themes are, because most good disinformation merchants know that images are more often exchanged than text. That’s why you see all those posters with text on

The finding the configurations part - we use Gephi to look at the network; botnets and distributors stand out like little flowers in a Gephi network. But we could use networkx to do the same thing. There are also a set of tools in OSOME that will help you examine relationships quickly.

Raw data is useful too - it’s where we start. But really, in social engineering, it’s the relationships that count.

1.2.2 Chasing botnets

I use bot sentinel and tools like it - ones like Hamilton68 monitor accounts from nation state actors (Russia, China etc - think embassy twitter feeds, RussiaToday etc), ones like Botsentinel monitor accounts active in earlier campaigns that might or might not be bots. The most valuable thing they give you is trends: what the recent chatter online is.

Bot detection is an art now. Once upon a time, it was as easy as “there are 100 accounts posting all the time, and they’re all posting the same text”, and finding them was basically “look for the Qanon hashtags”. Now it’s more subtle. There are some rules of thumb, like being suspicious of anything tweeting more than 100 times a day, but there’s more to it, and a bunch of tools to help.

1.3 Chasing an image

There are a few things you’re going to want to do with an image:

- Extract the text from it
- See where else it exists online
- Check to see if it’s been altered / is fake

Extracting text: You can usually extract text from images using OCR (optical character recognition). There are libraries like Tesseract that can be called from Python (as e.g. pytesseract), but they have mixed results. A more reliable way to do this is to use the OCR built into search engines to pull the text from each image: yandex.com appears to be best at this (although always check because OCR still doesn’t produce perfect results) but is Russian: if that’s an issue for you, bing.com image search does this too.

Seeing where else an image is online:

- Mostly you’ll be doing this by hand for new images, but a good first check is to see if an image (e.g. a photo) has been reused from an earlier event. Reverse image search from yandex.com and bing.com works well - tineye.com will call all the big image search engines for you (and you can laugh at some of the things they return…).

Checking for alterations: Bellingcat are the masters of online image forensics, and have a good guide to this (Bellingcat guide). Look at tools like FotoForensics.

1.4 Handling Video and Audio

1.4.1 Checking video

InVID_EU

1.4.2 Save an audio file from Facebook Messenger

The workaround is:

- Using Chrome browser (but NOT on mobile)
- Access facebook via m.facebook.com
- Then click on the messenger icon
- Go to the chat that has the audio
- Right mouseclick on the (...) at the end of the message and you'll have the option to "Save Audio As"

1.5 Searching through Facebook Groups

A lot of Covid19 disinformation is happening and/or moving at some point through facebook groups. We've been tracking some of these by hands whilst working out how to automate creating watchlists of groups, pages, accounts to check for new disinformation incidents forming before they hit the mainstream press.

Some academic references on this, focussed on antivax (one of the best-known and well-studied modern conspiracy theories)

- The online competition between pro- and anti-vaccination views
  - Hidden resilience and adaptive dynamics of the global online hate ecology
  - "New online ecology of adversarial aggregates: ISIS and beyond" with supplementary materials

Chapter 7a

Introduction

Disinformation analysis has changed a lot since 2016 when a search on #qanon, and some simple checks would find you botnets and a disinformation campaign. There are people who are good at disinformation data science (Eliot Alderson, Conspirador Norteno etc), and there's been a lot of academic money in this area recently. This section covers useful tricks, processes and tools.

Disinformation data science

Data science is a process. The team has a data science for disinformation response training series, including:

1. Setting up for disinformation activities: goals, ethics, groups, examples, practice
2. Disinformation basics: creators, outputs, mechanics, effects, feeds
3. Disinformation layers: from strategy/information ops down to tactics/artefacts/TTPs
4. Data collection: OSINT, platforms, user- and network- analysis level tools
5. Handling big data: APIs, cleaning, exploration, storage, automations
6. Communicating results: reporting routes, visualisation, tools, practice
7. Social text analysis: features, techniques, tools, generation methods
8. Image data analysis: tools, techniques, deepfakes/cheapfakes
9. Relationships as data: data-as-networks, features, tools
10. Using machine learning: extending your analysis with ML/AI
11. Acting (ethically): counters, coordinations and more ethics
12. Measuring effectiveness: measuring campaigns and counters

This was based on a university course, the aim of which was to take computer science students through the process, learning the basics of disinformation mechanics (there are many good papers on that), then walking through the data science processes that are particular and peculiar to tracking large disinformation campaigns across social media at speed.

The TL;DR is this is all about people. We need to think in terms of end-users, questions and problems.

There are different types of data scientist. One way to divide them up is by the time pressures that they work under:

- Strategic - months/weeks. Issue focussed (e.g. SJ's work on agriculture supply chains and covid). Good places to look for this type of work include Stanford Internet Observatory, UWashington, Shorenstein Center.
- Embedded data scientists - see these in dev teams. - days usually, sometimes on dev cycles (e.g. 3 weeks). Project focussed. Usually running hypotheses to support things like hypothesis-driven development and lean enterprise (pruning value trees etc). Look for these in the AI/ML-based disinformation tool companies.
- Data journalists - NY Times - days, sometimes overnight for a quick visualisation (long form journalism is more strategic). Good places to look include Bellingcat and DfrLab (these also do good strategic work).
- Tactical data scientists - hours/days. Incident-based. Basically the CTI League team, some of MLSEC, some of the crisis-mappers.

We could also divide data scientists by the things that they care about:

- Academics - long deadlines, care about papers and reputation
- Academics working on techniques - UIndiana
- Academics analysing actors and issues - DFRlab, strategic
- Government agencies / military - strategic
- Commercial interests -

Tactical data science

Working from the data we have instead is instructive and can teach us things about the disinformation environment, but that's not tactical data science. Most of the work so far hasn't been tactical. At speed, this becomes a threat intelligence nerd fight, that looks very similar to the other threat intelligence nerd fights: disinformation creators vs disinformation defenders.

A lot of what we do is detective work, where the algorithms and tools are there to assist us. This has a lot in common with data forensics, threat intelligence work and OSINT.

We need to think about data. Mostly we're dealing with data that's moving, at rest and static.

- Moving data: A lot of research places have social media listening - downloading all the social media messages etc around topics, hashtags etc of interest.
- Data at rest: this is the data we've grabbed during investigations, usually as part of finding more of a network and its effects. We're often actively analysing it, working out how we can affect the environment it's in.
- Static data: this data isn't going to change. Some of it is moving data that we've stored, and the environment it was in has been overtaken by events. It's of interest because it contains patterns to be mined, and could contain clues to later behaviours. Other static data is used to support investigations.

Where and how to look for examples

"tactical data science" is the work you do in the moment, chasing disinformation incidents and campaigns as they happen. There's a lot of literature out there on disinformation algorithm design, which is nice, useful in some circumstances (e.g. as a dayjob), but not helpful to people faced with "social media is happening, work out how to reduce harm". There's a lot there of the "there's a dataset, let's see what we can do with it" persuasion.

Places to look for ideas in this field include:

- Trained amateurs - sites like towards data science (lots of student projects), github, medium.
- Academics - known research groups, paper repositories, conference outputs
- Adjacent groups
- Student projects - yes, it's students, but they're usually supervised in latest techniques, keen to try them out online, and willing to write up their code.

Good search terms include "computational propaganda", "misinformation", "disinformation".

Places to look

- Action: check https://www.maltego.com/blog/mapping-visual-disinformation-campaigns-with-maltego-and-tineye/

    - Action: check https://www.secjuice.com/social-media-intelligence-socmint/
- Articles
- Google search “disinformation ‘data science’” - lot of posing. Found data science sites’ articles on disinformation.
    - DONE: continue on https://www.google.com/search?q=disinformation+%22data+science%22&oq=disinformation+%22data+science%22&aqs=chrome..69i57j0.3959j1j8&sourceid=chrome&ie=UTF-8
    - Action: mine https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU(2019)624278_EN.pdf for ideas / notes
- https://towardsdatascience.com/tagged/fake-news and searching the same site for “disinformation” and “misinformation”. There are several data science sites like this that might have articles of interest to us.
- Github searches
    - https://github.com/search?q=disinformation (119 repos)
    - https://github.com/search?q=misinformation (235 repos)

Disinformation Data Science Examples

Disinformation tracking:

- Fireeye looking at Iranian network: "Distinguished Impersonator" Information Operation That Previously Impersonated U.S. Politicians and Journalists on Social Media Leverages Fabricated U.S. Liberal Personas to Promote Iranian Interests and Network of Social Media Accounts Impersonates U.S. Political Candidates, Leverages U.S. and Israeli Media in Support of Iranian Interests
- Graphika tracking GRU network: From Russia With Blogs
- DFRlab on MyRan commercial operation: Facebook shut down commercial disinformation network based in Myanmar and Vietnam
- Vice reporting, unknown researcher on targetted deepfakes: Deepfakes by BJP in Indian Delhi Election Campaign

Data science

- Network detection and analysis (including botnets)
    - View of What types of COVID-19 conspiracies are populated by Twitter bots? - lots of good work in here (and reference from here)

    - The spread of low-credibility content by social bots - Filippo Menczer UIndiana work
- Account/activity analysis
    - Russian Fake Tweets Visualized (russian twitter dataset - NBC news)
    - https://secondaryinfektion.org/downloads/secondary-infektion-report.pdf (ACTION: mine this)
    - FAILED SURGE: Analyzing Beijing's Disinformation Campaign Surge On Twitter - analysing account creation timing
- Data exploration:
- - Sima's visualisations of Facebook disinfo data: Facebook's Coordinated Inauthentic Behavior - An OSINT Analysis
- Narrative tracking
    - Four experts investigate how the 5G coronavirus conspiracy theory began (not a great example)

Machine learning

- General
    - Graphika Labs Summer Retreat 2019: Innovating for the Future
- Temporal analysis
    - Fake News Classification via Anomaly Detection - time anomaly ideas
- GPT2 language generation / detection:
    - A Deep Learning Approach to Combating Misinformation* - student project? (GPT2 etc). Repo at stevenoluwaniyi/cognitive_computing: Cognitive computing application
    - Language Models and Fake News: the Democratization of Propaganda
- Article text classification (on the Kaggle dataset)
    - https://towardsdatascience.com/using-a-i-to-combat-fake-news-34f5a51907d
    - Getting Real with Fake News
    - How I built a simple Fake News detector on Amazon SageMaker
    - Getting Real with Fake News - student project?
    - Detecting Fake News With Deep Learning
    - How to build a recurrent neural network to detect fake news
    - Machine Learning tackles the Fake News problem
- Article text classification (not Kaggle)
    - Automatically Detect COVID-19 Misinformation - covid19 articles (interesting work on titles)

    - Machine Learning detects Fake News. (own 'dummy data')
    - Using USE (Universal Sentence Encoder) to Detect Fake News (not Kaggle)
    - I Built a Fake News Detector Using Natural Language Processing and Classification Models (theOnion vs notTheOnion)
    - Fake news or not? (also theOnion vs notTheOnion)
    - How I trained an AI to detect satire in under an hour - theOnion articles vs other scraped articles; used MachineBox
    - A Quick Guide to Fake News Detection on Social Media - 2017 survey of work
- Deepfake detection
    - Donald Trump is an Android. Or not real. According to this AI detector. text (Trump is a bot…)
    - Attention is All They Need: Combatting Social Media Information Operations With Neural Language Models - Fireeye on text generation and detection
- Tweet classification
    - Trawling Twitter for Trollish Tweets - Servian (uses 538 russian tweet dataset)

Tactical Disinformation Data Science Tasks

Disinformation data science tasks: The bigger things with machine learning (or augmented intelligence, e.g. humans and algorithms working together):

- Fact-checking: verifying that the content of an article, image, video etc doesn't contain disinformation. This is hard, and usually needs a team of fact-checkers, up-to-date knowledge etc.
- Source-checking: verifying that a source (e.g. a publisher, domain etc) doesn't distribute disinformation. This is why we label and track URLs. Several groups already publish labelled lists of domains.
- Source networks: fake news creators usually run multiple websites. Some work on "pink slime" exists.
- Detecting computational amplification
- Detecting fake accounts
- Detecting inauthentic account networks (including botnets)
    - Looking at patterns of account creation dates for popular messages
- Detecting, tracking and analysing narratives

General references

- Using Data Science to Detect Disinformation - useful concepts for large mixed teams
- Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature - useful backgrounder
- https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU(2019)624278_EN.pdf - useful look across European programs
- Ben Nimmo, Camille Francois, C Shawn Eib, Lea Ronzaud, Rodrigo Ferreira, Chris Hernon, Tim Kostelancik, "Sekondary Infektion", Graphika report 2020 - hand tracking through large-scale network; some good visualisation ideas

Test datasets

- Kaggle "getting real about fake news" Getting Real about Fake News - used a lot
- Twitter deleted 200,000 Russian troll tweets. Read them here. - NBC's Russian twitter dataset
- fivethirtyeight/russian-troll-tweets - 538's IRA dataset
- 538 dataset was from Salesforce's Social Studio tool ($1000/month) Editions & Pricing: Social Media Marketing

PS a lot of the examples are in Python and Pandas - you don't escape from learning these Python Data Science Handbook

This is where the data science comes in...

Narrative detection and analysis

Topic modelling

Chapter 8

Data science, data analysis, starts and ends with human beings. We can do beautiful analysis, but if we don't make it accessible to the people who need to take action from it, then we haven't done our job.

Let's talk about outputs. The ways we present the data we produce, and how we do that, including the forms/ formats some of the people we interact with are used to, what good visualisations in this space look like (and how to create them), and how to get those outputs to the right people.

Visualisations

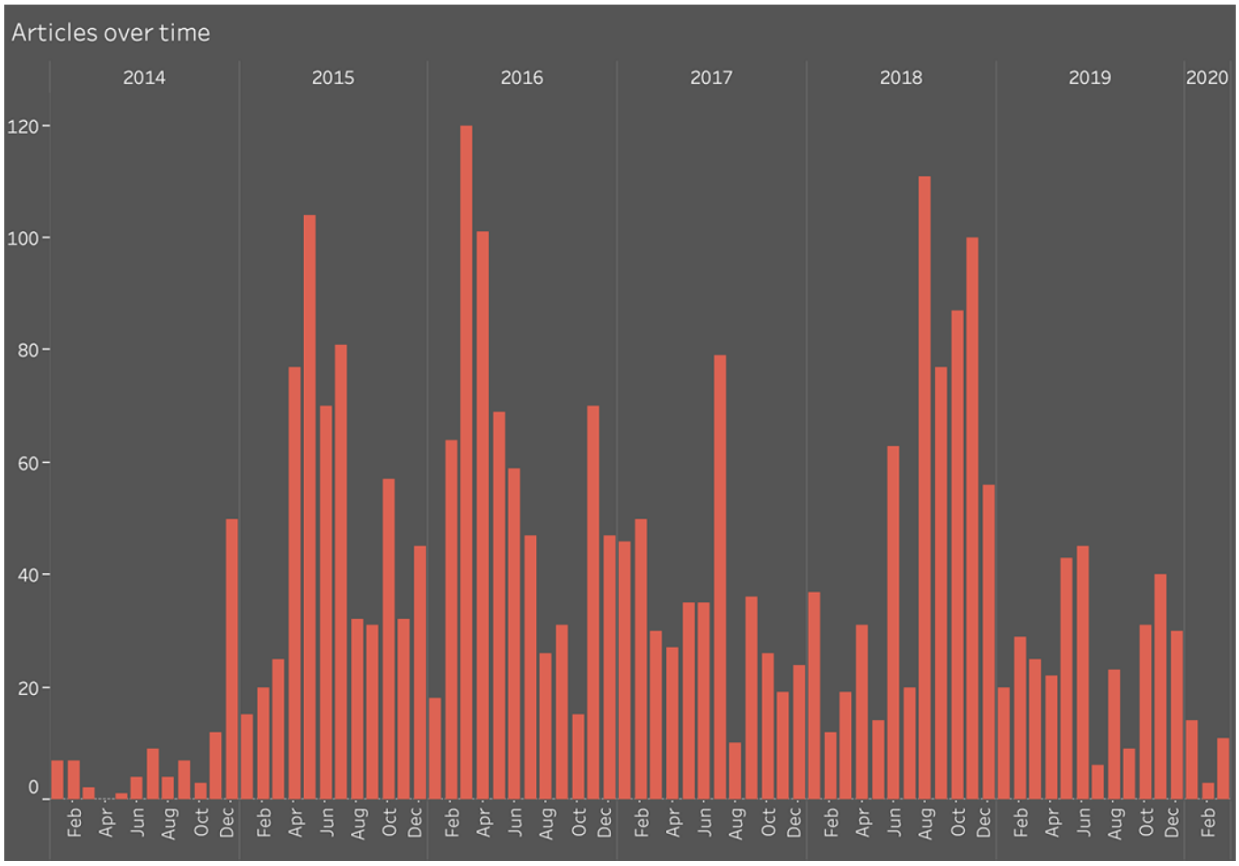
Eyeballing the data, looking at statistics, and examining machine learning outputs are good, but part of getting to know data, and explaining it to other people is being able to look at it visually. There's a lot of work on data visualisation (read "Storytelling with Data" to see it done well), so this section is looking at what disinformation people do with visuals.

Good places to look for what "that chart is" include

- All Charts - python visuals (most data scientists use Python)
- A Periodic Table of Visualization Methods - periodic table of visualisations

Understanding time series

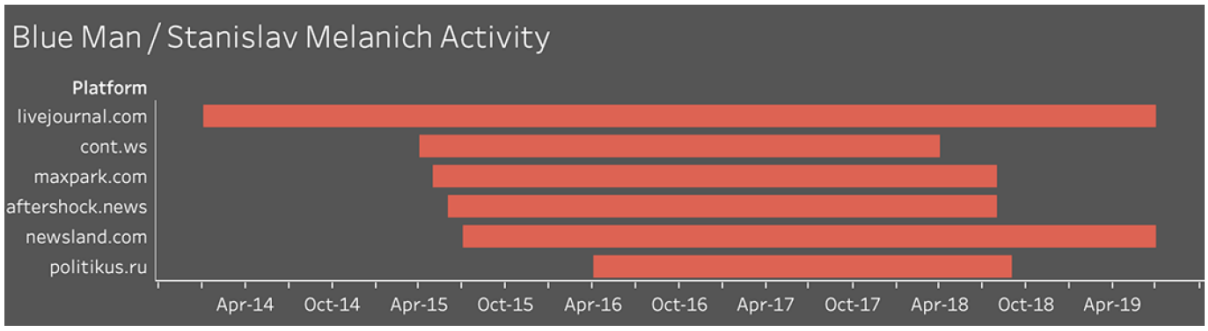
Disinformation operations happen over time, so time-based plots can be useful tools. The humble bargraph (or column plot) is really useful for this. Almost every visualisation tool has this as an option (e.g. https://matplotlib.org/3.2.1/api/_as_gen/matplotlib.pyplot.bar.html)



Secondary Infektion: number of articles posted over time.

(Sekondary Infektion report, 2020)

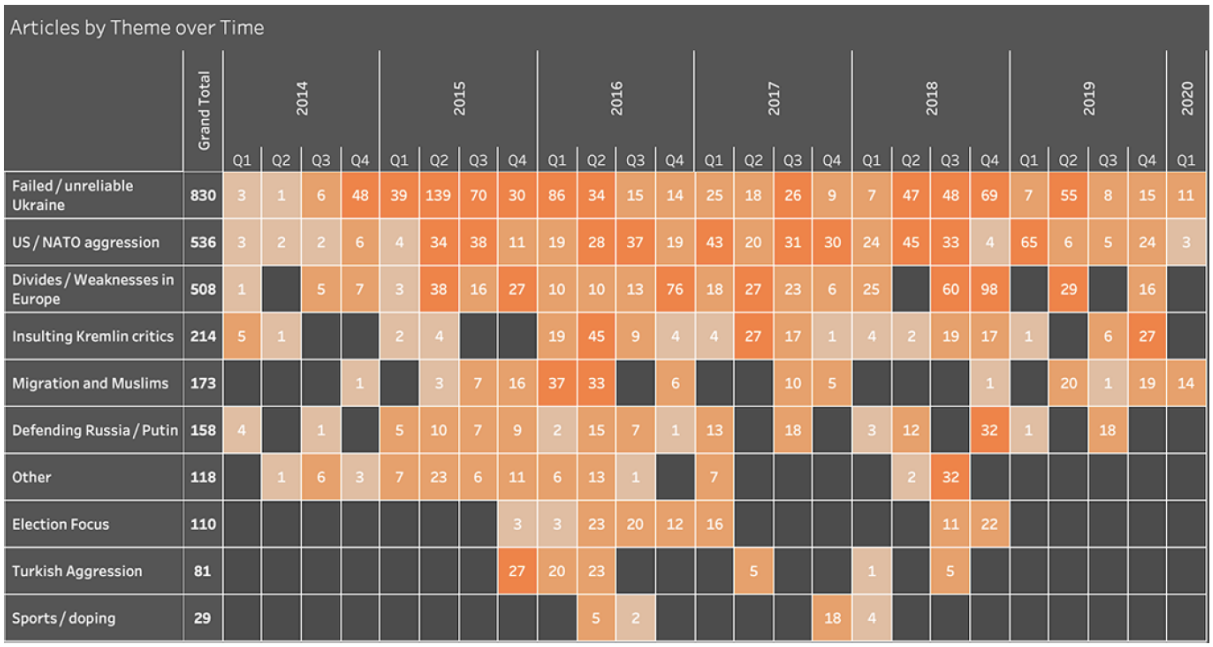
Bar graphs and line plots can be used for showing a range of entities over time.



The Blue Man's activity across different platforms, 2014-2019.

(Sekondary Infektion report, 2020)

If the value range is too large to show easily (e.g. there's a mix of very small and very large values that you can't easily plot on one axis), heatmaps might be more appropriate. https://python-graph-gallery.com/91-customize-seaborn-heatmap/

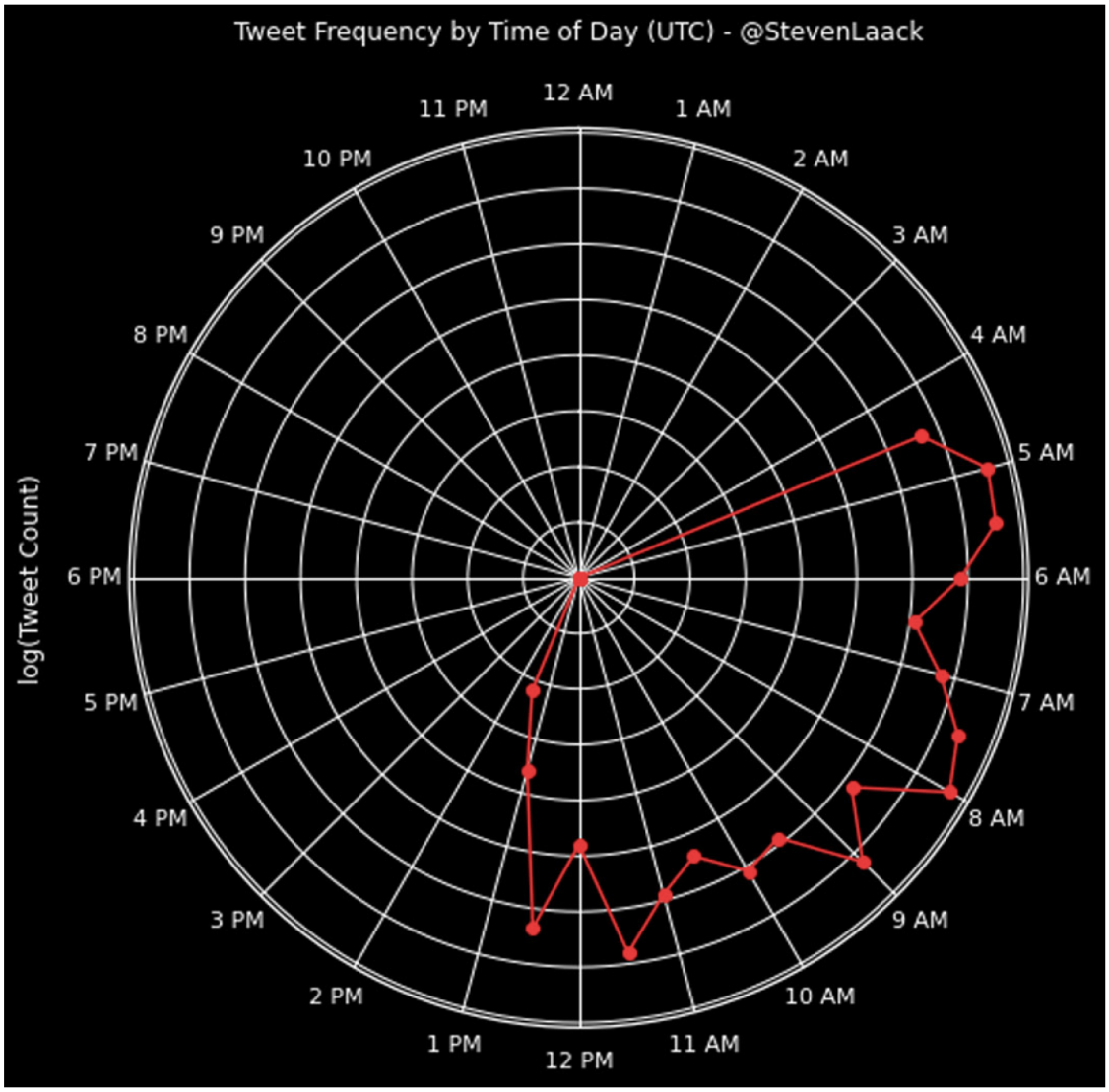


	Grand Total	2014				2015				2016				2017				2018				2019				2020
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
Failed / unreliable Ukraine	830	3	1	6	48	39	139	70	30	86	34	15	14	25	18	26	9	7	47	48	69	7	55	8	15	11
US / NATO aggression	536	3	2	2	6	4	34	38	11	19	28	37	19	43	20	31	30	24	45	33	4	65	6	5	24	3
Divides / Weaknesses in Europe	508	1		5	7	3	38	16	27	10	10	13	76	18	27	23	6	25		60	98		29		16	
Insulting Kremlin critics	214	5	1			2	4			19	45	9	4	4	27	17	1	4	2	19	17	1		6	27	
Migration and Muslims	173				1		3	7	16	37	33		6			10	5				1		20	1	19	14
Defending Russia / Putin	158	4		1		5	10	7	9	2	15	7	1	13		18		3	12		32	1		18		
Other	118		1	6	3	7	23	6	11	6	13	1		7					2	32						
Election Focus	110								3	3	23	20	12	16						11	22					
Turkish Aggression	81								27	20	23				5			1		5						
Sports / doping	29										5	2					18	4								

Secondary Infektion: main themes over time

(Sekondary Infektion report, 2020)

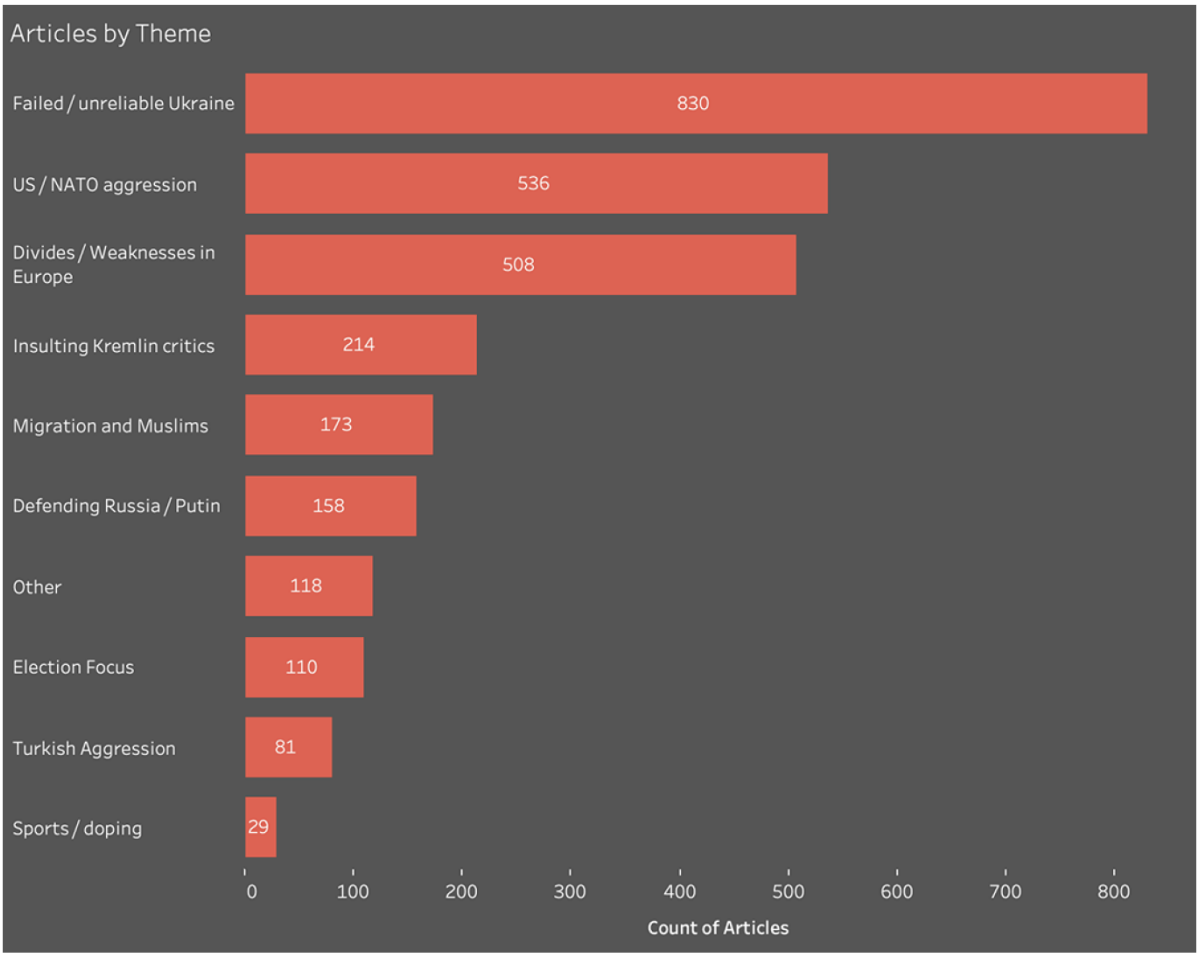
The use of spider plots for 24-hour data is good too, because they don't have a "start" or "end" time, making it easier to compare different diurnal patterns.



(Sekondary Infektion report, 2020)

Understanding relative sizes

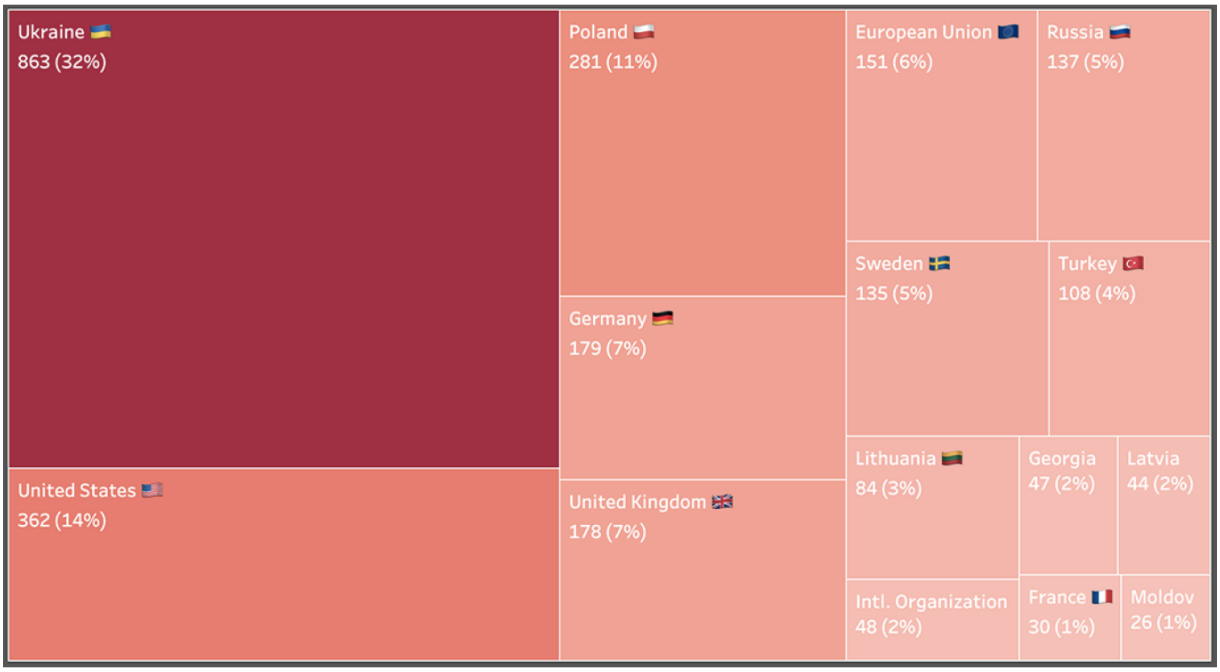
Bargraphs can do this too. http://python-graph-gallery.com/barplot/



Breakdown of Secondary Infektion articles by theme and number.

(Sekondary Infektion report, 2020)

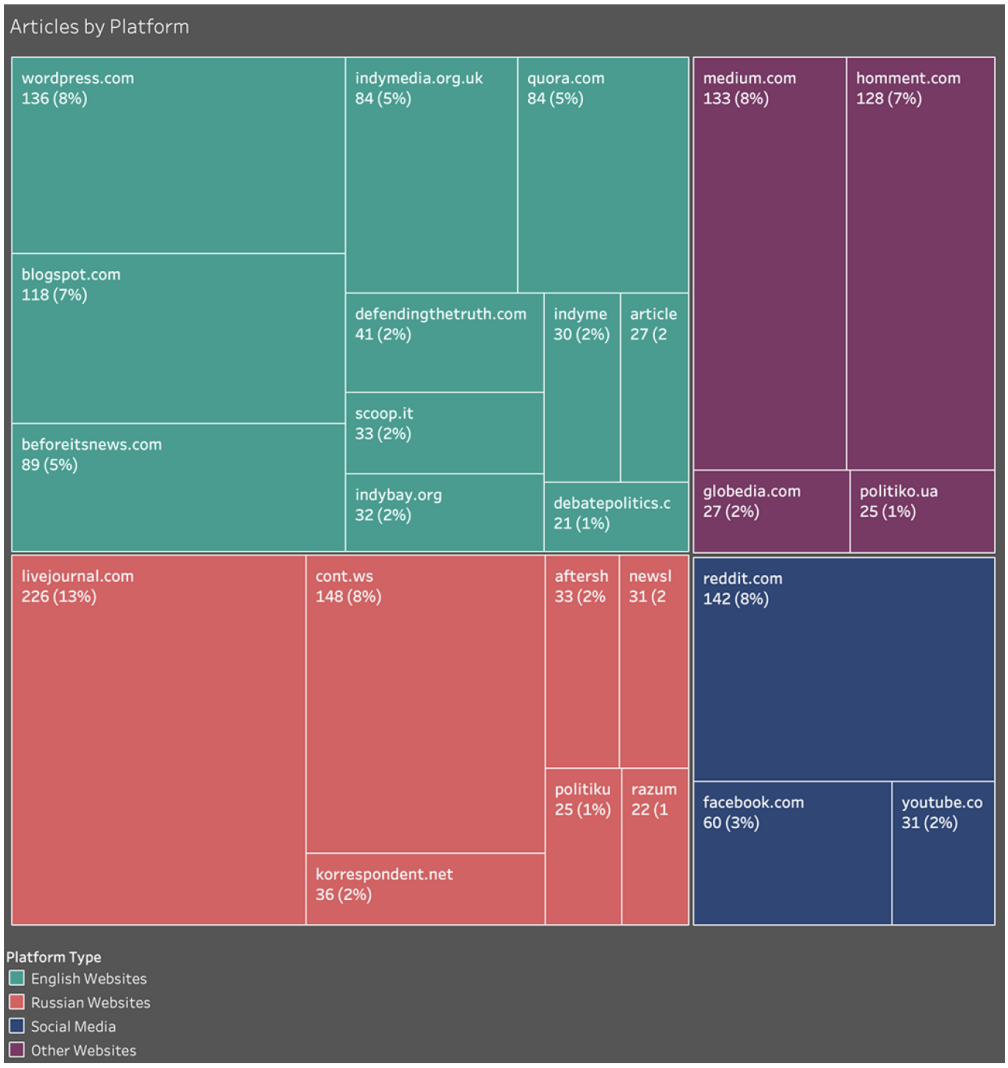
Treemaps show relative sizes as areas. https://python-graph-gallery.com/200-basic-treemap-with-python/



Countries mentioned or targeted by Secondary Infektion, total number of stories.

(Sekondary Infektion report, 2020)

Colours are an extra, useful, dimension on most plots.

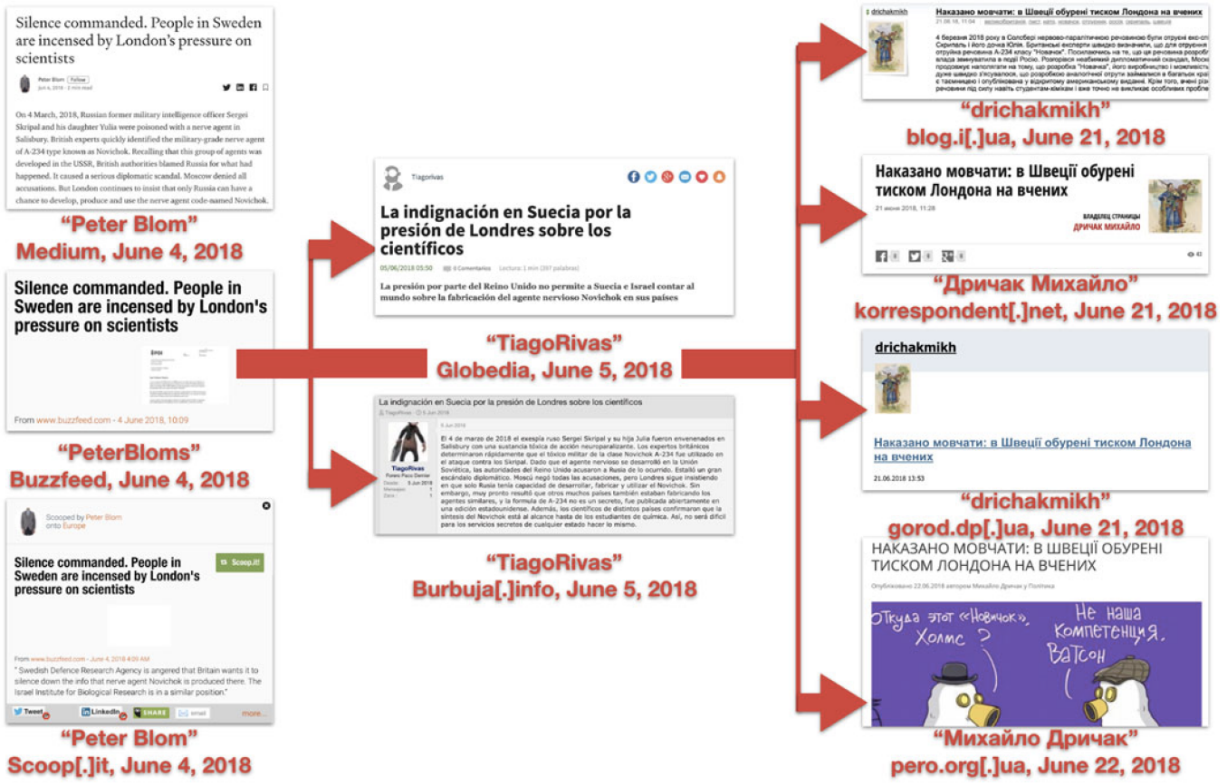


Use of platforms: where Secondary Infektion posted the most.[179]

(Sekondary Infektion report, 2020)

Understanding connections

Really simple graphics (think powerpoint) can help explain the connections between objects.

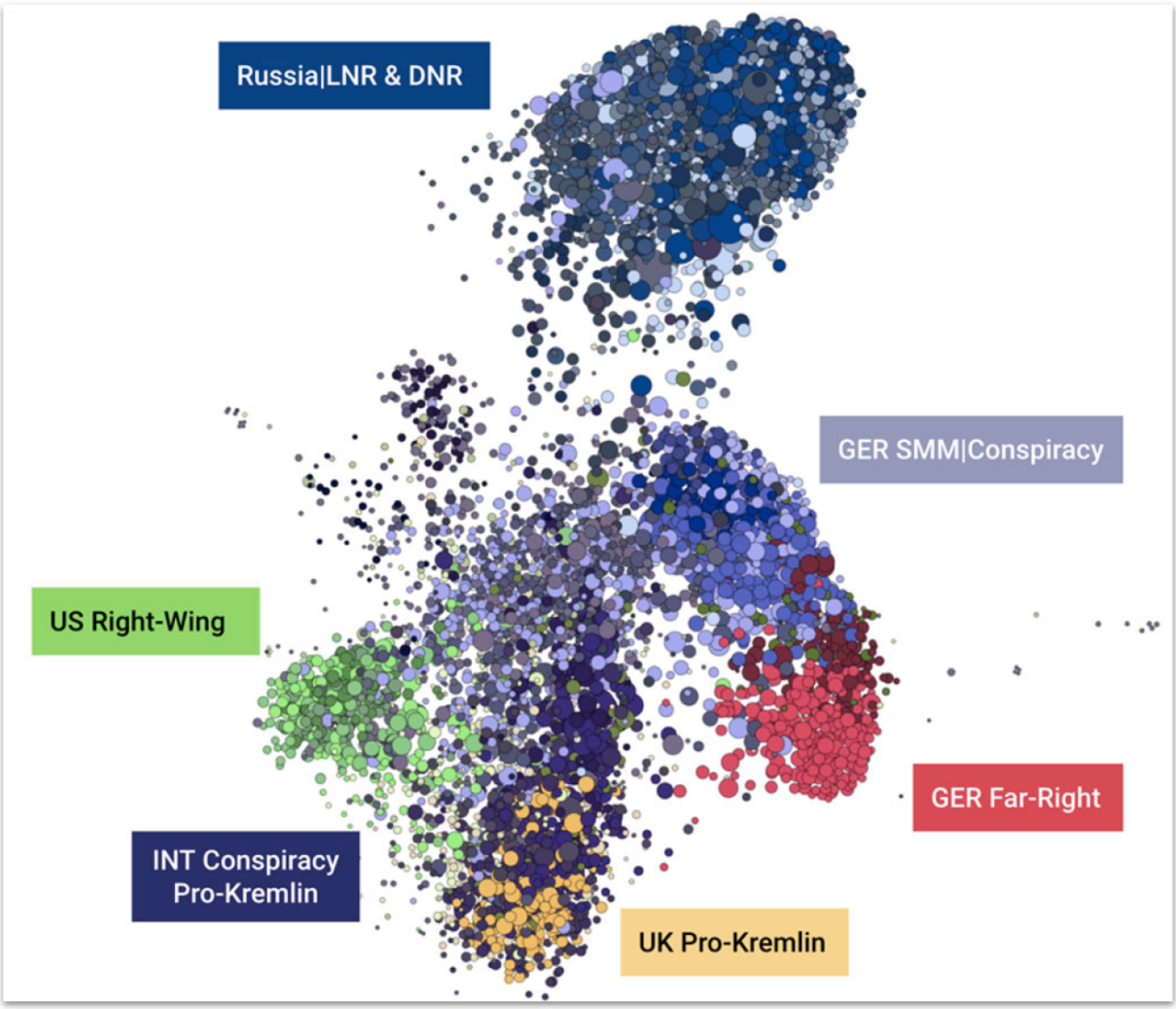


A simplified diagram of the spread of the article by assets commanded by Secondary|Infektion, from the English-language persona "Peter Blom" or "Peter Bloms" (left), through the Spanish-language persona "Tiago Rivas" (center) and the Ukrainian-language persona Михайло Дричак ("Mikhailo

(Sekondary Infektion report, 2020: nice use of arrows)

Simply gridding out pages or accounts with the same visuals or information can be really powerful if you're describing a network.

Graph diagrams show a large number of nodes and the connections between them - the "snurfball" images that we sometimes show to explain where the influencers are in an incident. Tools like Gephi produce these, with a little work (and liberal use of things like the Force Atlas 2 algorithm to make the network structure easier to see). Graphika produces "network maps" - one explanation is "The circles represent individual Twitter accounts. The volume of the circles represent influence by following, while the colours represent political ideologies.". This also looks like graph diagrams.



Graphika network map of the Secondary Infektion Twitter assets' followers and the followers of significant amplifiers, mapped April 2020.

Others

Some visualisations are hard to classify - is this a network diagram or the output from a dimension reduction algorithm? (dimension reduction = a type of machine learning algorithm that takes a set of objects that exist in many dimensions, and flattens it so it's easy to see - usually as a two-dimensional plot).

Specialist text analysis: look at things like Scattertext
https://towardsdatascience.com/hkprotest-visualizing-state-troll-tweets-from-chinas-disinformation-campaign-1dc4bcab437d

Chapter 9

The point of real-time disinformation tracking is to be able to do something about it. Our basic actions include:

- Direct action.
- Asking someone directly connected to us to take action
- Reporting to someone not directly connected to us, so they can investigate and decide whether to take action.

Direct action: there are many small things that a team could do to disrupt a disinformation incident. These include:

- Flooding a disinformation hashtag or group with alternative information (be careful with this because if the original intent was confusion, you might be adding to it) etc

Asking someone connected to us to take action

- Reporting a suspicious domain to registrars. If we do this, it's on us to gather information to help them - e.g. screenshots of selling bleach 'cures' etc etc

Reporting to someone not directly connected

- This is most likely with the large social media platforms. We're going to find bots and botnets; we won't be able to remove them ourselves, we will be able to report them to platforms. It'll help if we have that reporting mechanism set up ahead of time.

1.1 Reporting

1.1.1 Reporting inside the League

If you know which organisation you need, use the /list_orgs and /list_contacts [org] slack command to find the person you need. More generally, look at the channels guide in the League handbook to see the right channel to report an incident or component to.

1.1.2 Reporting to law enforcement from the League

You can open an LE escalation ticket using the /lenew command

1.1.3 Reporting to platforms

Reporting to social media

- Reddit: https://www.reddit.com/r/redditsecurity/
- Twitter: report-twitter-impersonation and twitter-rules
- Facebook: How to Report Things on Facebook
- Linkedin: Reporting Inaccurate Information on Another Member's Profile
- Instagram: https://help.instagram.com/1735798276553028
- YouTube: https://support.google.com/youtube/answer/2802027
- Google: is going to take some digging Avoid and report Google scams - Google Help

Pinterest

- Fast: https://help.pinterest.com/en/article/report-something-on-pinterest
- Slower: report on https://help.pinterest.com/en/contact?page=about_you_page - you'll need a Pinterest account to do this from.
    - Choice is porn, violence, hate speech, self harm, harassment/ exposed private information, spam; currently going with either hate speech, violence or harassment as appropriate.
    - Has an image filesize limit of 2MB
- community guidelines are https://policy.pinterest.com/en-gb/community-guidelines

Reporting websites

If you've found a website or ring of websites, teams you can report it to (with supporting notes) include registrars and the lists used by adtech and other sites to check the types of sites that they're passing money through.

Site lists:

- Global disinformation index
- Media bias fact check

1.2 Direct Action

1.3 References

Chapter 10

1.1 HIVE

We use Hive to manage our list of incidents, and links from them to the other objects and data connected to incident responses. Check Hive and search for the incident name. All incidents will have the tag “disinformation” and word “Incident” in the title, which should help with searching.

1.1.1 (Adding an object workflow to a Hive Incident - don’t use this yet)

Adding a new workflow to a case:

1. Assume the current Case ID is (A).
2. Create a new Case (B) selecting the workflow Case Template you wish to add to Case (A).
3. Open Case (A) and click "merge".
4. Select "By Number" and add Case ID (B).

1.2 MISP

https://bbb.secin.lu/b/ale-q6v-ecn <-- Recorded MISP Training for COVID courtesy the CIRCL folks

1.2.1 Adding an object (tweet etc) to MISP by hand

- Go to MISP
  - Click on the incident ID in the list of events.
- Click on “Add Object” in the left-side column
  - Misc -> microblog for twitter or Facebook posts
  - Fill out the details
  - Click submit
  - Repeat for more objects

- Now you can start playing with the grey bar at the bottom of the event description, and toggle things like the timeline on and off.

Object types we're most likely to need are:

Object	Misp	Hive equivalent
Facebook group	misc:facebook-group	url
Facebook page	misc:facebook-page	url
Facebook account	misc:facebook-account	url
Facebook post	misc:facebook-post	url
Twitter account	misc:twitter-account	url
Twitter list	misc:twitter-list	url
Twitter post	misc:twitter-post (was misc:microblog)	url
Blogsite	network:url	url
Blog account	misc:user-account	url
Blogpost	misc:blog	url
Reddit group (subreddit)	misc:reddit-subreddit	url
Reddit account	misc:reddit-account	url
Reddit post	misc:reddit-post	url
Reddit post comment	misc:reddit-comment	url
YouTube Channel	misc:youtube-channel	url
YouTube Video	misc:youtube-video	url
YouTube Playlist	misc:youtube-playlist	url
YouTube Comment	misc:youtube-comment	url

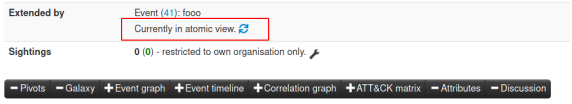
Website address	network:url	url
Hashtag	ADD NEW	hashtag
Instant message	misc:instant-message	
Instant message group	misc:instant-message-group	
Narrative	misc:narrative	
Image	file:image	
Meme	file:meme-image	
Individual	misc:person	
Event (e.g. protest)	misc:scheduled-event	
Location	misc:geolocation	

Other objects we might need include:

Object	Misp	Hive equivalent
	misc:course-of-action	
	network:email	
	file:forged-document	
	file:leaked-document	
	misc:legal-entity	
	misc:news-agency	
	misc:organization	
	misc:scheduled-event	

	misc:short-message-service
	network:shortened-link
	misc:user-account

1.2.2 Adding an object to MISP via Slack bot

- Slack bots can quickly create and append an object to an event.
- Each bot attempts to modify the MISP event directly. If it lacks permission it will instead create a MISP event extension. Click the icon shown below to switch to extended mode to see the extended event objects appended into the main event.
- 

1.2.2.1 Twitter Posts

There's a Slackbot in #4-disinformation that can upload a Twitter post to a MISP event. The bot works like this /misp_twitter $MISP_event_id $post_id

It accepts either a Twitter Status ID or a Twitter post URL as arguments for $post_id

- In the #disinformation channel use the following command to add a Twitter post to the CTI League MISP
  - /misp_twitter <misp event id> <twitter post URL or twitter post ID>
  - Example: /misp_twitter 34 https://twitter.com/NASA/status/1259960728951365633?s=20

1.2.2.2 BuiltWith Tags

- In the #disinformation channel use the following command to add a Twitter post to the CTI League MISP
  - /misp_builtwith <misp event id> <url or domain name>
  - Example: /misp_builtwith 34 newyorkcityguns.com

1.3 DKAN

DKAN is a data warehouse tool - it's where we store large datasets and their descriptions,for analysts to use.

1.4 Gephi

1.4.1 Viewing networks with Gephi

This is a manual process with instructions created from Andy Patel's video at

https://www.youtube.com/watch?time_continue=17&v=AqlT0khVuZA

- Get Gephi from https://gephi.org/users/download/ - install it.
- Start Gephi.
- Click on top menu>file>"import spreadsheet". Grab User_user_graph.csv - use all defaults
- Top menu: Go to data laboratory, "copy data to another column", click 'id', click okay.
- Go to overview. RHS: Run modularity algorithm, using defaults
- RHS: Run average weighted degree algorithm
- LHS: Click color icon, then partition, modularity class. Open palette, generate, unclick "limit number of colors", preset=intense, generate, okay
- LHS: Select "tt", ranking, weighted degree, set minsize=0.2, choose 3rd spline, apply
- LHS: Layout: OpenOrd, run. Then forceatlas2, run. Try stronger gravity, and scaling=200
- Top menu: Preview - select "black background", click "refresh". Click "Reset zoom"

Gephi has an API - these tasks could be automated.

1.5 Slack bots

We use slack bots to push artefacts to MISP.

we can now add the following object to a MISP event using the following slash commands

/misp_reddit_account

/misp_reddit_post

/misp_reddit_comment

/misp_reddit_subreddit

If we want new ones - we can build them, and [redacted] wrote a handy how-to guide: https://vvx7.io/posts/2020/05/misp-slack-bot/

If we want new MISP object types, here's how to do that too:

1. Create the new object folder
    1. Git clone https://github.com/MISP/misp-objects
    2. Go into repo folder objects. It contains a subfolder for every misp object type
    3. Copy one of the existing object folders; rename the copy to the new object you want
    4. Go into the new object's folder. You'll find one file in here: definition.json. Open it for editing
2. Set basic data
    1. Get a new UUID from https://www.uuidgenerator.net/ - replace "uuid" in definition.json with this new one
    2. Set "version" to 1
    3. Set "name" to the same as the new folder name (nb use "-" not "_")
    4. Set "description" to something descriptive
    5. "Meta-category" is usually "misc"
3. Set attributes. Go through attributes. For each one, set:
    1. "Description": something descriptive
    2. "Misp-attribute": see https://www.circl.lu/doc/misp/categories-and-types/. You'll probably use "text" a lot. The difference between url and link? url isn't trusted; link is trusted (this signals whether something is safe to click on).
    3. "Ui-priority": just leave this as default (1 is always okay)
4. These attributes aren't mandatory, but are useful
    1. "Multiple": set this to "true" if you allow multiple of this attribute (e.g. hashtags)

    2. "disable_correlation": true, - stops MISP trying to correlate this attribute - set this on things like language to stop MISP from wasting time
    3. "to_ids" - makes exportable via api - set to false as needed (most attributes don't need it)
5. Set the list of attributes that an object must have one of to exist
    1. List these in "requiredOneOf"
6. Check the new object is valid
    1. Run validate_all.sh
    2. Run jq_all_the_things.sh
7. Push your change back to the MISP objects repo (or to Roger for sanity-checking)

1.6 Python scripts

We use python a lot (just look at the github repo…). Here are some useful resources:

- Learn python the hard way
- ACTION: [redacted] add notes on python and data science - [redacted]-level friendly

---

1.7 Other Tools

We've mentioned a bunch of tools above.

Some basic tools:

- Most data scientists use Python and Jupyter notebooks. You'll see a lot of these - the basic Anaconda install comes with most of the things we use https://www.anaconda.com/distribution/
- Data gathering:
    - Reaper https://github.com/ScriptSmith/reaper https://github.com/ScriptSmith/socialreaper https://reaper.social/ - scrapes Facebook, Twitter, Reddit, Youtube, Pinterest, Tumblr APIs
- Network analysis and visualisation: there are many tools for this.
    - Gephi is a good standalone tool https://gephi.org/users/install/

    - Networkx is a useful python library
- URL analysis
    - Builtwith.com
- Image analysis
    - Reverse image search: tineye.com, Bellingcat guide
    - Image search: bing.com, yandex.com
    - Image text extraction: bing.com, yandex.com
- Data storage / Threat Intelligence tools
    - DKAN https://getdkan.org/
    - MISP https://www.misp-project.org/

Disinformation-specific tools:

- Indiana University has a set of tools at https://osome.iuni.iu.edu/tools/
    - Botometer: check bot score for a twitter account and friends https://botometer.iuni.iu.edu/#!/
    - Hoaxy: check rumour spread (uses Gephi) https://botometer.iuni.iu.edu/#!/
    - Botslayer https://osome.iuni.iu.edu/tools/botslayer/
- Bellingcat made a list of useful tools
    - Bellingcat's really big tools list - worth reading if you need a specific OSINT tool

Chapter 11

1.2 Bedtime Readings

1.2.1 Books

If you really want to get into how we got here, the history of information operations, what disinformation and propaganda are etc, these books were recommended by the team:

- [redacted]s 2018 book stack - dated, but some good classics in here
- Thomas Rid's "Active Measures"
- PW Singer and Emerson Brooking's "Like War"
- Zeynap Tufeki's "Twitter and Tear Gas" (free version)
- Verification handbook: handbook, investigative reporting
- Verification Handbook: homepage

1.2.2 Articles

- Unpacking China's Viral Propaganda War
- Prevalence of Low-Credibility Information on Twitter During the COVID-19 Outbreak (5 pages)
- Media Manipulation and Disinformation Online (106 pages)
- Facebook's Coordinated Inauthentic Behavior - An OSINT Analysis
- Naval Post Graduate - Disinformation (many)
- Hate multiverse spreads malicious COVID-19 content online beyond individual platform control (9 pages)
- From Russia with Blogs (26 pages)
- The COVID-19 Social Media Infodemic (18 pages)
- We've Just Seen the First Use of Deepfakes in an Indian Election Campaign
- Facebook shut down commercial disinformation network based in Myanmar and Vietnam
- Facebook April 2020 Coordinated Inauthentic Behavior Report (26 pages)
    - Iran's Broadcaster: Inauthentic Behavior (46 pages)
    - Facebook's VDARE Takedown (18 pages)
    - Facebook Downs Inauthentic Cluster Inspired by QAnon (19 pages)

- (Bellingcat) Uncovering A Pro-Chinese Government Information Operation On Twitter and Facebook: Analysis Of The #MilesGuo Bot Network
- Unmaking Democracy: How Corporate Influence Is Eroding Democratic Governance (Harvard International Review) - 4 May 2020 (6min read)
- Conspiracy Theory Handbook (12 Pages)
    - Google Drive Location
- What if we've all been primed? (6 pages)
- (Bellingcat) Investigate TikTok Like a Pro (15min read)

1.2.3 Podcasts and videos

- Motherboard's Cyber Podcast Episode with Thomas Rid about Active Measures and implications for modern disinformation
- Lawfare's Arbiters of Truth podcast series about disinformation
    - Especially this episode with Camille Francoise specifically about COVID-19 disinfo and the ABCs of Disinfo
- vOPCDE #2 - Discussion: Disinformation about Disinformation ([redacted])
- **1.2.4 People to Follow**

Disinformation data science:

- Conspirador Norteno and Dr ZQ: @conspirator0 @ZellaQuixote
    - always a great example on bot tracking
    - went looking at reopen etc https://twitter.com/conspirator0/status/1252374902121721859?s=19
    - Tools: https://makeadverbsgreatagain.org/allegedly and python/jupyter with libraries pandas, tweepy, bokeh, cytoscape
    - Looking at a botnet: https://twitter.com/conspirator0/status/1265829648056954881?s=20
- Andy Patel: @r0zetta
    - Infosec and misinformation data scientist
    - Tools: e.g. using TFIDF plus Louvain clustering to analyse twitter https://twitter.com/r0zetta/status/1230786764413030400 https://twitter-clustering.web.app/ (https://github.com/r0zetta/meta_embedding_clustering)
    - https://blog.f-secure.com/author/andrew-patelf-secure-com/page/2/
- Elliot Alderson: @fs0c131y (fs0c131y.com)
    - Infosec and misinformation data scientist

Disinformation tracking:

- Erin Gallagher: @3r1nG
- @josh_emerson
- Kate Starbird: @katestarbird

1.2.5 Examples of disinformation tracking

- "Distinguished Impersonator" Information Operation That Previously Impersonated U.S. Politicians and Journalists on Social Media Leverages Fabricated U.S. Liberal Personas to Promote Iranian Interests
- From Russia With Blogs
- Facebook shut down commercial disinformation network based in Myanmar and Vietnam
- Facebook's Coordinated Inauthentic Behavior - An OSINT Analysis

Disinfo data science (short investigations)

- https://onezero.medium.com/facebook-groups-and-youtube-enabled-viral-spread-of-plandemic-misinformation-f1a279335e8c

Images and disinformation

- Deepfakes by BJP in Indian Delhi Election Campaign

Disinformation Counters

- Training end-users about disinformation
  - https://getbadnews.com/#intro - game to train people on how disinformation works
  - CrashCourse media literacy videos

Chapter 3a

1.4 Tracking Covid19 Disinformation outside the USA

America and the UK are original masters at disinformation campaigns (both for their work from second world war onwards, but also for the internal propaganda work so successfully picked up later by e.g. China). Russia, China, Iran are all biggies right now in online disinfo aimed at other countries, but there are also countries whose internal (aimed at their own population) disinformation campaigns have been masterful (Venezuela) or unsubtle but effective (Philippines). There are other countries where the use of disinformation is just kinda background normal politics, but generally internal and local (e.g. Nigeria). A very subjective top 10 list would be: USA, China, Russia, Iran, UK, Saudi Arabia, Pakistan, India, Venezuela, Philippines.

Things to think about: who

- How is a country involved?
  - Disinformation customer / originator
  - Disinformation target
  - Disinformation producer / factory
- What type of disinformation?
  - Geopolitics / Nation State propaganda: country A to country B/C/etc
  - Politics / propaganda: country A to own population
  - Grifting: individuals to population (usually for money)
  - Power: groups to population (recruiting, actions etc)

Things to think about: what

- Localisation:
  - Local tech use (including social media)
  - Local power structures
  - Local concerns
  - Languages
  - Communication style
  - Local idioms (e.g. “cockroaches”)
- Globalisation

    - Common themes: politics, grifters, 5g, antivax etc

Places to look for non-USA disinformation:

- Disinformation repositories
    - https://euvsdisinfo.eu/disinformation-cases/ - Russia disinfo on EU
    - https://medium.com/dfrlab - world disinfo
    - https://comprop.oii.ox.ac.uk/ - nationstate actors
        - Specifically The Global Disinformation Order and case studies

https://www.newsguardtech.com/covid-19-resources/ - c19 domains for several countries

- Hive cases, MISP events etc
    - E.g. reopen starting in Australia, moving to Canada etc

---

1.5 Non-Covid19 disinformation and where to send it

It's almost certain that in the course of looking for Covid-19 related disinformation, we're going to find disinformation on other topics. While our mandate is specifically Covid-19 related, there are other, area-specific organizations to which we can report disinformation.

- Right-wing extremism/hate speech: Southern Poverty Law Center
- Voter suppression attempts:
    - On social media, report the post to the platform using their reporting mechanisms
    - You can also report the issue to the U.S. Department of Justice
- Anti-GLBTQ+: Gay and Lesbian Alliance Against Defamation

1.3 Covid19 disinformation references

Disinformation

- https://tomnikkola.com/prime/
- "The Dark Arts of Disinformation Through a Historical Lens"

- "The Kremlin's Disinformation Playbook goes to Beijing/
- "Anti-Lockdown Protests Originated With Tight-Knit Group Who Share Bigger Goal: Trump 2020"
- "NATO STRATCOMCOE considers 'Disinformation in Asia'"
- "Activists fight COVID-19 disinformation in the Caucasus"
- "Anatomy of a disinformation campaign: The coup that never was"
- "Recognizing Disinformation during the Covid-19 Pandemic"
- "Disarming Disinformation"
- "Tech giants recalled by MPs over lack of 'adequate answers' on disinformation"
- "'The country is in a state of trauma': COVID-19 has made the US a breeding ground for propaganda and a goldmine for foreign spies"
- "House Democrats' coronavirus bill earmarks $1 million to study 'disinformation'"
- "Disinformation 'whack-a-mole' doesn't work on social media"
- "Belarus, Moldova, and Ukraine: COVID-19 disinformation in Eastern Europe"
- "How TikTok could be a player in election disinformation"
- "EU tackles coronavirus disinformation, seeks regulatory framework for Facebook, other social-media companies"
- "EU demands tech giants hand over data on virus disinformation"
- "Wexton seeks study of COVID-19 disinformation, misinformation"
- "Uncovering A Pro-Chinese Government Information Operation On Twitter and Facebook: Analysis Of The #MilesGuo Bot Network" Also briefly about Covid19 disinformation network

Covid19 Narratives

- "COVID: Top 10 current conspiracy theories"

Covid19 disinformation around the world

- India:
  - "India & COVID-19: Misinformation and the Downside of Social Media" - whatsapp, fake cures, SM responsible for curation, strong messaging from Modi
  - "How 300 Indian scientists are fighting fake news about COVID-19"
- Italy:
  - "Italian MP amplifies debunked COVID-19 conspiracy theories on the floor of Parliament"
- China:

    - "China in coronavirus propaganda push as US ties worsen" - hero story
- Africa:
    - "Coronavirus: What misinformation has spread in Africa?"
    - "The other COVID-19 pandemic: Fake news"
    - Nigeria Centre for Disease Control - countering
- Venezuela:
    - "The coronavirus infodemic in Latin America will cost lives"
- Ecuador:
    - "Another virus is causing major damage in Ecuador. It's called fake news" - targetted by bot farms from neibouring countries
    - Información chequeada sobre el Coronavirus - Latam countering (case lists, in Spanish)