

README: About The BigBook

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. We have modified the chapters from their original form by removing CTI-specific notes and adding in other notes that might help other teams who are engaging in similar work. We've released them in the hope they can be useful to other teams running disinformation defense.

This book is intended for use by groups setting up their own disinformation monitoring.

Book Structure

- Introduction
 - [Chapter 0 About this book](#)
- What are you trying to do?
 - [Chapter 1 Disinformation](#)
 - [Chapter 2 Disinformation Models](#)
 - [Chapter 3 Influence techniques](#)
- Getting organised
 - [Chapter 4 Looking after yourself](#)
 - [Chapter 5 The Disinformation Response Team](#)
 - [Chapter 6 Disinformation Response Networks](#)
- Building evidence
 - [Chapter 7 Monitoring](#)
 - [Chapter 8 Incident Response](#)
 - [Chapter 9 Detective work: Asking the Right Questions](#)
 - [Chapter 10 OSINT: Investigating Artifacts](#)
 - [Chapter 11 Data Science: Investigating Big Data](#)
 - [Chapter 12 Machine Learning: Pattern Hunting](#)
- Taking Action
 - [Chapter 13 Creating a Situation Picture](#)
 - [Chapter 14 Sharing Information](#)
 - [Chapter 15 Countermeasures](#)
- Useful References
 - [Chapter 16 Tools](#)
 - [Chapter 17 General References](#)
 - [Annex A: Covid-related disinformation](#)
 - [Annex B: Election-related disinformation](#)

The original BigBook was a set of GoogleDocs. We [moved it to GitBook](#), but baulked at the fee to keep printing PDFs. So now it's here.

C0 About This Book

TL;DR The Fast Route	1
Introduction	1
Where This Book Came From	1
Distributed Disinformation Defense	2
Contributors	2
CogSecCollab	3
The CTI League	3
Covid19Activation and Covid19Disinformation	4
Fellow Travelers	4
Crismappers, Old and New	4
Helpful info	4
Glossary	5
Other places to look for information	6

TL;DR The Fast Route

Every chapter in this book contains detailed instructions on how we do the things we do, why we do them, and ways to look for more information about that. When the sticky stuff hits the rotating thing, there isn't time to read through all that, so each chapter will also start with a "TL;DR" section, containing the basics needed to get set up and running.

Introduction

Where This Book Came From

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. We have modified the chapters from their original form by removing CTI-specific notes and adding in other notes that might help other teams who are engaging in similar work.

We've written a lot of text during our Covid19 disinformation response, and we're realizing 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. When we finish, it should contain notes on:

- * What disinformation is, for threat intelligence (TI) people
- * How to run distributed disinformation TI teams
- * Suggested data standards for that
- * Suggested tools for that
- * Resources that could help with disinformation threat sharing and response

Distributed Disinformation Defense

This book is about the things we need to build, test and document to create truly distributed disinformation defenses. That includes the tools, techniques and resources for threat sharing and response as part of threat-informed defence, how to use those resources in communities, and how other communities (e.g. MLsec and other parts of information security) might help with this.

The first big idea is that disinformation campaigns are a large-scale distributed, asymmetric threat, and we need a large-scale, collective, distributed, asymmetric response to them. This includes:

- joining together all the individual responses,
- connecting community alerts to responders,
- making it easy,
- making it possible to build groups for free,
- connecting new groups to the existing information security response system

The next big idea is that near-real-time distributed disinformation response is easier if you can share information and data quickly, and the MISP threat intelligence platform, lightly extended, lets you do that.

This is systems work, and most systems contain people, processes, algorithms, data, and insights. Of all these, the most important part is the people. Steven from the CS-ISAO pointed out that what we've accidentally built is a [Security Operations Center \(SOC\)](#) for disinformation - over time, we'll adapt the nomenclature in this book to more closely match that.

Contributors

We are community. We have been part of many communities on this journey, helped form communities and organizations, 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](#) is a volunteer community of academics, researchers, technologists, students, journalists, and engineers, all working on the intersection of infosec and misinformation.

Its 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, CogSecCollab is concentrating on building the tools and techniques that groups need to counter disinformation campaigns.

CogSecCollab officially 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.

- Before 2019, we were raising awareness that there was an intersection between infosec and disinformation / information operations.
- In 2019, we built the AM!TT Framework, an infosec-based description framework for influence operations, and started models of influence operation countermeasures.
- In 2020, we're building tools, processes and community to support community-based responses to counter disinformation campaigns.

CogSecCollab also convenes the MISP disinformation sharing community, which is focused on influence operations. The community is listed in the MISP communities page on <https://www.misp-project.org/>.

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. Formed in April 2020, the CTI League identifies, analyzes and neutralizes all threats. At this most sensitive time, the CTI League 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. However, there are some things that we do a little differently, which is how we ended up writing a book.

Covid19Activation and Covid19Disinformation

We work alongside other teams. Covid19Activation was started by TEDx fellows to track Covid19 information and disinformation, and has worked tirelessly on this.

Covid19Disinformation was started jointly by CogSecCollab and Andy Carvin from the Atlantic Council's DFRlab team as a community disinformation tracking team (Andy is one of the original crismappers). Covid19Disinformation fell fallow as the CTI League team built up, but provided many of the baseline processes for the CTI League's Disinformation team.

Fellow Travelers

Tool adaptations and trials, specifically the MISP and AMITT adaptations, were the work not just of the CogSecCollab and CTI League teams. We were joined in this work by teams from NATO, DHS, MITRE, RRM Canada, EEAS, MISP, CIRCL and Belgium, specifically in the [MISP disinfo sharing community](#). These are our wider community.

Crisismappers, Old and New

Ten years ago, some of the CogSecCollab leads worked in and with the community data teams that responded to the Haiti 2010 earthquake, ran globally-distributed but locally-focused teams, and created processes and tools for what became the crismapping movement. Crismappers changed the ways that disaster response and development agencies managed both data and the way that they worked with people on the ground. We see many parallels between that work, and the work of creating and sustaining community-supported disinformation response.

Helpful info

Glossary

We all come from different disciplines: words like "campaign" have different meanings to a military, adtech or 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 working 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 color)
- 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, mislabeled 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 behavior around it, like flooding a hashtag. That activity usually lasts for a short period of time because the narratives, artifacts, 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 components of incidents. Each incident might have multiple narratives involved or just one, but there's usually an identifiable narrative somewhere in there. You can use narratives to see if there are related incidents that have already been tracked or dealt with. Narratives, like incidents, have lifetimes. Some narratives appear as a result of a world or local event (or anticipated event), and are only useful while that event is in peoples' minds.
- Artifact: Artifacts are the objects that you can 'see' connected to a disinformation incident or campaign. Artifacts are the text, images, videos, user accounts, groups, hashtags, etc. that you use to get a picture of an incident or campaign.

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

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 are separate books being written about that. This BigBook is the practical one.

We've added lists at the end of this document, under References, that will point you to books and papers about disinformation, other teams doing this, sources of data, tools, etc. In addition, CogSecCollab is collecting information in its [documentation repo](#), which was used to seed this BigBook.

Chapter 1 Disinformation

TL;DR: Looking for Disinformation	1
Defining Disinformation	2
Good introductions to disinformation	3
Online Influence and its abuses	3
What People Do Online	3
With people comes value	4
With value comes targetting	4
With value comes abuse	5
Disinformation as a Digital Harm	5
Reading	6
Disinformation from the Creators' POV: Intent	6
Where disinformation comes from	6
Disinformation for Money	7
Disinformation for GeoPolitics	8
Disinformation for Politics	9
Disinformation for Power	9
Disinformation for Business	9
Disinformation for Lols and Attention	10
Satire and Conspiracies	10
Reading	11
Disinformation from the Defender POV	11
What disinformation targets	11
Big, Fast, Weird: why disinformation is getting harder to track	12
Further Reading	13

TL;DR: Looking for Disinformation

You're looking for deliberate promotion of false, misleading or misattributed information in your team's specific area of interest. This might be motivated by geopolitics (countries interfering with the beliefs and sentiments of each others' populations), money (selling webpage views, clicks, t-shirts, 'cures' etc), politics, power (e.g. adding people to extremist 'in-groups'), or LOLs. You're not here to stop debate, however odious it is - you are here to help reduce online harms.

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 working on practical disinformation response, try not to get sucked into that very large rabbit hole: 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 or emotions 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 or emotions in a large number of people”) and
- mechanisms (“focus on creation, propagation, consumption of misinformation online”).

We are countering deliberate creation and propagation of false information online - whether that falsehood is in the information itself, who it purports to come from, the groups set up to output it, etc. We care a lot when it’s designed to change beliefs or emotions in large numbers of people.

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 targeting 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.

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.

What People Do Online

- [Social networks](#) (examples include MySpace, Facebook, and LinkedIn)
- [Micro-blogging websites](#) (examples include twitter and StumbleUpon)
- [Blogging and Forums websites](#) (examples include WordPress, tumblr, and LIVEJOURNAL)
- [Pictures and Video-Sharing websites](#) (examples include YouTube, flickr, and Flikster)
- [Music websites](#) (examples include Pandora, lost.fm, and iLike)
- [Online Commerce websites](#) (examples include eBay, amazon.com, and Epinions)
- [Dating Network websites](#) (examples include match.com, eHarmony, and chemistry.com)
- [Geo Social Network websites](#) (examples include foursquare, urbanspoon, and tripadvisor)
- [News and Media websites](#) (example include the LA Times, CNN, and New York Times)

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.

With people comes value

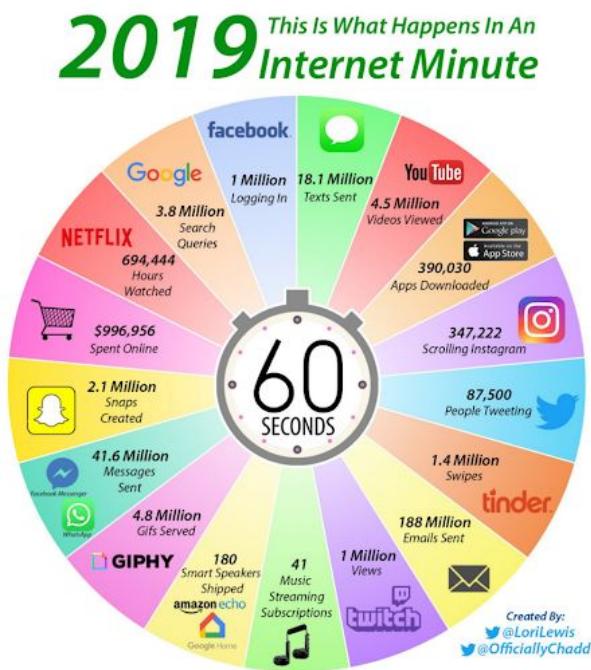


Figure: Internet Minute, 2019

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.

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

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](#)".

- Physical harm: e.g. bodily injury, damage to physical assets (hardware, infrastructure, etc).
- Psychological harm: e.g. depression, anxiety from cyber bullying, cyber stalking etc
- Economic harm: financial loss, e.g. from data breach, cybercrime etc
- Reputational harm: e.g. Organization: loss of consumers; Individual: disruption of personal life; Country: damaged trade negotiations.
- Cultural harm: increase in social disruption, e.g. [misinformation](<https://www.washingtonpost.com/politics/2020/02/21/how-misinformation-n-whatsapp-led-deathly-mob-lynching-india/>) creating real-world violence.
- Political harm: e.g. disruption in political process, government services from e.g. internet shutdown, botnets influencing votes

When we assess whether to respond to a disinformation incident, we use a set of criteria that include an estimate of the risk, defined in terms of the potential harm caused by the incident without response, for example in the League, we're specifically looking for, and trying to reduce, medical harms (another criterion is whether other teams are already responding to reduce the incident's potential harm). The use of the harms framework above is important because in 2020, social media companies, politicians etc shifted their definition of 'bad' on social media from immediate calls for violence to the idea of digital harms where the effects might be delayed.

Reading

Internet history:

- [An Internet History Timeline: From the 1960s to Now](#)
- <https://www.slideshare.net/debbylatina/internet-history-190741201>
- We Are Social: [Global digital report 2019](#) - internet size

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, [Tracing Disinformation Trajectories from the 2010 Deepwater Horizon Oil Spill](#), 2016

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.

The big idea here is that the Internet makes everyone a nation-state actor now.

Disinformation for Money



FACTCHECK.ORG®

HOME ARTICLES ▾ ASK A QUESTION ▾ DONATE ARCHIVES ▾ ABOUT US ▾ SEARCH MORE ▾

[DONATE](#)

DEBUNKING FALSE STORIES

Fake Coronavirus Cures, Part 1: MMS is Industrial Bleach

By [Saranac Hale Spencer](#)

Posted on February 11, 2020



Quick Take

Online posts have claimed to reveal various “cures” for the novel coronavirus. Some are benign, like eating boiled garlic, while others are potentially dangerous, like drinking chlorine dioxide, an industrial bleach. Neither will cure the virus.

Covid19-related sales included t-shirts, cures, blood, views, clicks

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); inadvertently [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 deep fakes - 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 teens, and not all from Macedonia, but there were some Macedonian villages that were centres for disinformation production), 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.

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; levers they can pull on to influence other nations. These are resources available in pursuit of national objectives, usually referred to as 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.

Disinformation for Politics

Disinformation for Power

SIGN UP

A screenshot of a Facebook post from the group 'The Real 3%ers Idaho'. The post was made 'about a month ago' and has a profile picture of the group. The text of the post reads:

ATTN Ada County BUSINESS OWNERS in Boise and Surrounding areas:
We have credible intel tonight that Antifa and other groups are planning a riot tonight in the Boise area. Their plan is to destroy private property in the city and continue to residential areas. We are calling on all business owners to contact us if you are concerned for your business and your private property immediately. We are here to protect you, your private business, and have teams on the ground standing by.

Below the post, there are engagement metrics: 202 likes, 87 comments, and 410 shares.

2020 social media message

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.

Disinformation for Business

There's a business equivalent to the DIME model (Breuer&Perlman):

- Business deals and strategic partnerships
- PR and advertising
- Mergers and acquisitions
- R&D and capital investments

All of these things can be attacked using disinformation campaigns.

Disinformation for Lols and Attention

Have you seen this shark?

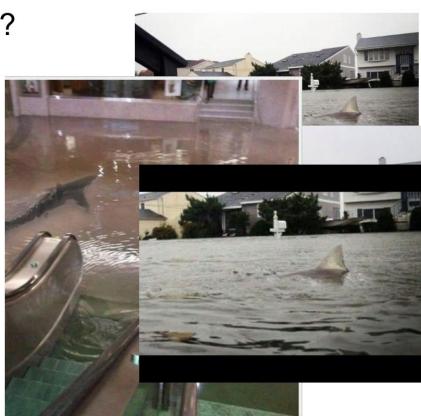
Believe it or not, this is a shark on the freeway in Houston, Texas.
#HurricaneHarvey



11:00 PM - 27 Aug 2017

85,254 Retweets 143,636 Likes

8.9K 13.8K 144K



2

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. Crisis mappers see it and sigh, then ask the original poster to remove it please because it's messing up the online response (crisis mappers 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. Crismappers 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, crismappers 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.

Satire and Conspiracies

Generally, we ignore disinformation for LOLs unless it's flooding an important hashtag or group (the social media equivalent of a DDOS). The only caveat on that is that satire and conspiracies are sometimes used as a gateway into more worrying narratives and groups, and might merit attention for that.

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](#)
- Also check <https://www.psywar.org/articles>

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 nation state. 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 targeting 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 targeted disinformation. Individuals (BillGates, Fauci) have had targeted disinformation campaigns around them; some campaigns directly targeted 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 voices 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, Weird: 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: Cliff Stoll's "The Cuckoo's Egg" and Mike van Putte's "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.

Further Reading

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:

- [SJ's 2018 book stack - dated, but some good classics in here](<http://overcognition.com/2018/12/10/misinformation-readings/>)
- Thomas Rid's "[Active Measures] (<https://us.macmillan.com/books/9780374287269>)"
- PW Singer and Emerson Brooking's "[Like War] (<https://www.likeyarbook.com/>)"

- Zeynep Tufekci's "[Twitter and Tear Gas](<https://www.twitterandteargas.org/downloads/twitter-and-tear-gas-by-zeynep-tufekci.pdf>)" (free version)
- Verification handbook: <http://verificationhandbook.com/>, specifically the chapter on investigative reporting <http://verificationhandbook.com/book2/chapter1.php>

Understanding infosec:

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

Chapter 2 Disinformation Models

TL;DR Disinformation Models	1
Deconstructing Disinformation	1
Disinformation Models	2
Cognitive Security	2
Disinformation Layers	3
Disinformation Pyramid	3
Actors in the Pyramid	4
Disinformation Objects	5
STIX and extensions	5
References	7
Disinformation Typographies	7
Disinformation TTPs: Tactics, Techniques, Procedures	8
AMITT Framework	8
Example: Double Diamond	9
Example: Plandemic	10
Reading	11
Social Media Object Set	12
Narrative Models	12

TL;DR Disinformation Models

We're using adapted information security standards to describe disinformation incidents, so we can share them with a large number of responders.

An incident is a coordinated sets of activities, over a relatively-short timespan, usually with an individual or team behind it. We describe these in terms of narratives (the storylines in the incident), TTPs (techniques used), and incident objects (actors, tools etc). We use STIX to describe most incident objects, AMITT to describe techniques, and text to describe narratives.

Deconstructing Disinformation

Part of CogSecCollab's work has been to model and populate the processes, tools, and countermeasures needed to respond in a distributed collaborative way to disinformation at scale. This chapter looks at some of the underlying models we use to understand and describe disinformation.

Disinformation Models

There are overlaps between disinformation, information security, machine learning, and military competition short of armed conflict. Whilst ideas from all of these appear feature in this book, the models and processes that we base most of our work on come from information security.

The big ideas here are:

- information security and disinformation defence are so similar that we can use the same tools for them both.
- If we have a common description language, we can share information about disinformation incidents in real time
- If we describe the moves disinformation creators use, we can mitigate or block those moves

This chapter covers mapping misinformation attack and defense into existing infosec frameworks, and how that mapping allows us to plan misinformation defenses and counters, assess tools and mechanisms, and handle the types of large-scale adaptive threats that machine learning applied to information security, MLsec, makes possible. This should help misinformation response move past the intelligence and passive defense stages of the SANS sliding scale, to architectural and more active defenses.

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. For example, if you look at definitions of information warfare, they include electronic warfare, military deception, operation security, computer network operations, and psyops, where CNO is what we typically think of as infosec, and psyops as including disinformation.

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.

We do have a name clash here, but not a severe one: some groups use "cognitive security" to mean using AI in information security defence, but that's not often heard in MLsec communities.

In here, when we talk about cognitive security, we're talking about one of the three main layers of information security:

- 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 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, Cognitive Security.

Disinformation Layers

Disinformation Pyramid



We use the disinformation pyramid to show how we connect information operations, threat intelligence, osint research and disinformation data science.

- 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). Information operations work is often at this level.
- 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. OSINT research and threat intelligence usually happens on this level.
- 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.
- Artifacts: 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. Data scientists usually start here.

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.

Actors in the Pyramid

For power-motivated disinformation, we have three main groups of people: the creators of misinformation ('attackers'), the people trying to counter them ('defenders'), and the targets of the misinformation ('populations'). Typically, attackers start at the top of the pyramid and work their way down. Defenders are at the bottom and work their way up.

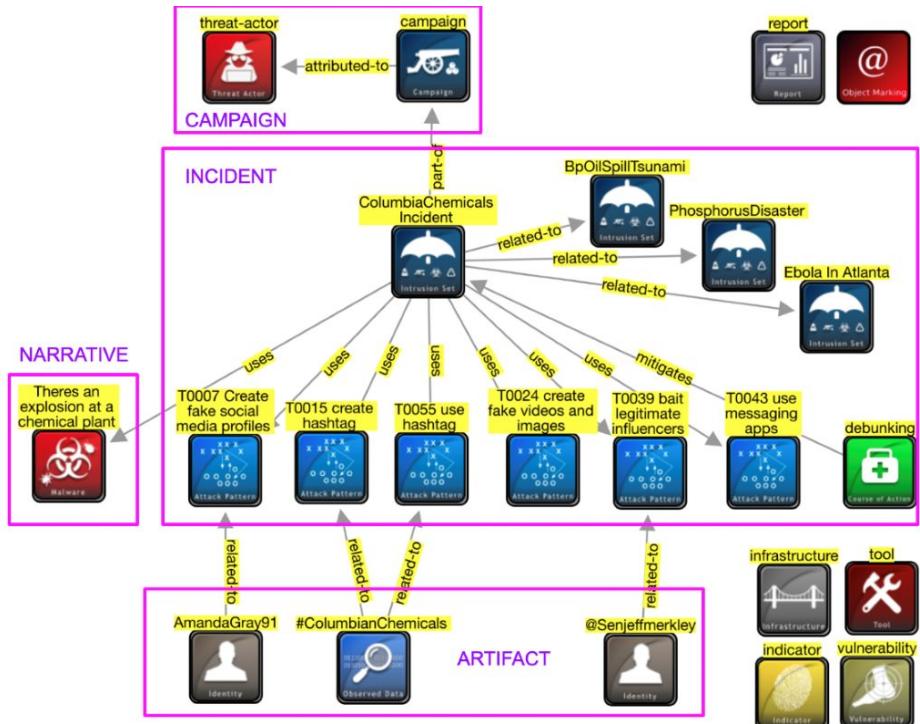
- Red. Attackers create incidents (e.g. Macrongate), which often form part of longer-term campaigns (e.g. destabilize French politics). Human communication is generally at the level of stories, or narration: we tell each other stories about the world, as gists or memes. And to tell these stories, we need artifacts: the users, tweets, images etc that are visible in each attack. Attackers have a goal they want to accomplish and design a misinformation campaign to achieve that goal. They manufacture one or more incidents, each incident has its own narrative which is told through a series of artifacts. Those artifacts can be posts, tweets, stories, deep fakes, etc. As attackers move down the pyramid, more work must be done. A single campaign can have thousands of artifacts transmitted by tens-of-thousands of accounts.

- Blue. Whilst the attacker sees the whole of the pyramid from the top down, the defender usually sees it from the bottom up, working back from artifacts to understand incidents and campaigns, unless they're lucky enough to have good insider information or intelligence. Most current misinformation work is at the artifact level, although there has been narrative (story) level work happening recently. By contrast, defenders start at the bottom of the pyramid. They see an artifact, and then another, and at some point, they may be able to tie all of these artifacts into a cohesive narrative. Eventually several of these narratives can be tied to distinct incidents and with enough investigation and perhaps a little attribution, a campaign can be discovered. This is definitely an "uphill" climb. Defenders will never uncover every artifact and are likely to miss numerous narratives and incidents because they simply don't have access to the communities and platforms where they present. Even with access, they may never get around to analyzing the information or even recognize it as linked to a campaign.
- Non-team. This is cognitive security, so there are many other actors in the pyramid, including people unwittingly sharing disinformation, or being the targets of disinformation narratives.

When you look at that pyramid, those layers aren't just about information - they're also about action, and understanding how to tie together both attack and defence activities from different layers. Information operations tends to be at the top layers, data science at the bottom: people doing this work need to be able to talk to and work with each other.

Disinformation Objects

STIX and extensions



STIX diagram for Columbia Chemicals

STIX is a data standard used to share information between threat intelligence organisations like ISACs. It's a rich language that describes threat objects and the relationships between them, is extensible, used by existing threat intelligence sharing communities (ISACs, ISAOs etc) so we'd be patching into an existing sharing system. It's also supported by and integrates well with existing community-supported, open-source tools.

Misinformation STIX	Description	Level	Infosec STIX
Report	communication to other responders	Communication	Report
Campaign	Longer attacks (Russia's interference in the 2016 US elections is a "campaign")	Strategy	Campaign
Incident	Shorter-duration attacks, often part of a campaign	Strategy	Intrusion Set
Course of Action	Response	Strategy	Course of Action
Identity	Actor (individual, group, organisation etc): creator, responder, target, useful idiot etc.	Strategy	Identity
Threat actor	Incident creator	Strategy	Threat Actor
Attack pattern	Technique used in incident (see framework for examples)	TTP	Attack pattern
Narrative	Malicious narrative (story, meme)	TTP	Malware
Tool	bot software, APIs, marketing tools	TTP	Tool
Observed Data	artefacts like messages, user accounts, etc	Artifact	Observed Data
Indicator	posting rates, follow rates etc	Artifact	Indicator
Vulnerability	Cognitive biases, community structural weakness etc	Vulnerability	Vulnerability

Disinformation version of STIX

STIX translates well for disinformation use. We added two objects to STIX for disinformation: incident, and narrative, and didn't need to change anything else. We use custom objects to represent these fields and be OpenCTI compliant.

For example, an Information Sharing and Analysis Center (ISAC) might share information about attacks against an industry via STIX/TAXII. Companies that are members of the ISAC then collect this (and other) information in a threat intelligence platform, then feed this information onto their security devices. They might also skip the threat intelligence platform and feed information from the ISAC directly to their security devices.

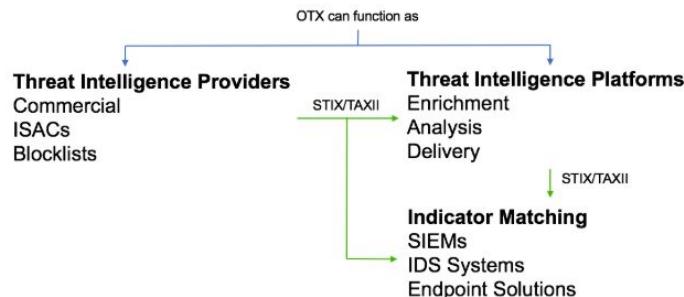


Image from https://stixproject.github.io/about/STIX_Whitepaper_v1.1.pdf

AMITT is now available as a STIX 2.0 bundle, from https://github.com/cogsec-collaborative/amitt_cti. When STIX 2.1 delivers an incident object we'll migrate to that.

References

STIX

- <https://www.alienvault.com/blogs/security-essentials/otx-is-now-a-free-stix-taxii-server>
- <https://pukhraj.me/2019/01/27/what-does-a-national-cyber-shield-look-like/#more-861>
- Einstein program
- STIX... and OpenC2
- https://stixproject.github.io/about/STIX_Whitepaper_v1.1.pdf
- <https://threatconnect.com/stix-taxii/>
- <https://www.crowdstrike.com/blog/indicators-attack-vs-indicators-compromise/>
- https://oasis-open.github.io/cti-documentation/stix/intro?_ga=2.135668339.378020639.1559740731-781460544.1559740731

Disinformation Typographies

STIX gives us objects, e.g. threat actor, but doesn't give a standardised way to describe the type of each actor, e.g. nationstate threat, for-profit threat, etc. We're working on that, with NATO, based on DFRlab's Dichotomies of Disinformation.

References

- <https://github.com/DFRLab/Dichotomies-of-Disinformation#appendix-a-codebook>

Disinformation TTPs: Tactics, Techniques, Procedures

AMITT Framework

misinformation-tactics		Analysis		Initial								Measure Effectiveness										
Strategic Planning (4 items)		Objective Planning (2 items)		Develop People (3 items)		Develop Networks (2 items)		Microtargeting (3 items)		Develop Content (10 items)		Channel Selection (10 items)		Pump Priming (8 items)		Exposure (10 items)		Go Physical (2 items)		Persistence (2 items)		
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														

AMITT TTP Framework, as seen in MISP

One of the disinformation objects that gives us a lot of information is the TTPs (techniques, tactics, procedures). In 2019, the Credibility Coalition MisinfosecWG team built a disinformation equivalent to the ATT&CK framework: the AM!TT (Adversarial Misinformation and Influence Tactics and Techniques) TTP framework, incorporating components from existing infosec standards, misinformation models, psyops, and marketing models (e.g. sales funnels), and designed using a wide range of example incidents, ranging from nationstate to small-group in-country operations. AM!TT's language and style is adopted from the MITRE ATT&CK framework, and its form is designed so we can use all the tools available for ATT&CK on it. CogSecCollab continues to be involved in the evolution and maintenance of AM!TT, including the use of subtechniques in the model.

AMITT is designed to give responders better ways to rapidly describe, understand, communicate, and counter misinformation-based incidents. We use the AMITT framework to break each disinformation incident down into its component TTPs, and to design and use

TTP-level countermeasures. It's 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> - in there, you can view a populated framework, where you can click on a technique and get details about what it is, who uses it, and which counters are available for it.

Every AMITT component has a unique id (e.g. T0018 Paid targeted ads). The framework is read left-to-right in time, with the entities to the left typically (but not necessarily) happening earlier in an incident. Its components include:

- Phases (not shown): higher-level groupings of tactic stages, created so we could check we didn't miss anything. The phases are separated into left-of-boom (planning, preparation) and right-of-boom (execution, evaluation), to represent activities before (left) and after (right) an incident is visible to the general public. The tactics below each phase belong to that phase.
- Tactics (top row): stages that someone running a misinformation incident are likely to use
- Techniques (all other rows): activities that an incident creator might use at each stage. The techniques below each tactic belong to that tactic. An example of a technique is T0010: Cultivate ignorant agents. This describes pulling in unwilling agents - through hiring them, or co-opting through emotion, agenda, sympathy (eg. conspiracy theorists are often ignorant agents). The technique doesn't define how to achieve this. There are many ways to hire or co-opt individuals, each potentially requiring its own counter.
- Tasks (not shown): things that need to be done at each stage. Tasks are things you do, techniques are how you do them.

To use AMITT, list and share the components you see in your incident: AMITT is now built into the MISP tool, making this easy to do. 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.

MITRE ATT&CK 2.0 is in the works and it refactors high level capabilities into implementations. It's a direction we'd like to see with AM!TT and something our group will continue to work toward.

Example: Double Diamond

We saw the use of T0010 (Cultivate ignorant agents) in the Double Deceit incident. This technique was used by EBLA, registered as an NGO in Ghana, doing NGO charity work on the

ground in Ghana (stationary to students). Problem is that that it's actually a Russian front. They hired local youth to post on social media. These people were not aware that they were part of a Russian troll farm. They also attempted to co-opt legitimate US influencers, to retweet their messages. Content was not political, and this appeared to be audience building.

We believe this is associated with the IRA. It appears to have been a failure.

Example: Plandemic

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

Plandemic TTPs in AMITT

The AMITT framework was built to be practical. We need to be able to translate our findings into an actionable story.

Plandemic is a debunked conspiracy theory video which makes some false claims about the nature of COVID-19. Despite high production quality the self reported cost to produce the film was less than \$2000. Zach Vorhies, an individual associated with QAnon, claims to be the social media marketer behind the viral success of the video. NYT reported his GoFundMe campaign titled "Help me amplify Pharma Whistleblower Judy Mikovits."

We can map out this small, but successful, operation in the AMITT framework to help us understand what capabilities the actor has and potentially how they're resourced. As with ATT&CK, we can start building an understanding of actors' capabilities over time.

Use fake experts

- Type: Technique
- Name: Use fake experts
- Id: T0045
- Summary: Use the fake experts that were set up in T0009. Pseudo-experts are disposable assets that often appear once and then disappear. Give "credibility" to misinformation. Take advantage of credential bias
- Tactic: TA08

Technique T0045, used in Plandemic

Plandemic exploited credential bias, and relied heavily on AMITT technique T0045: use fake experts. The "expert" here was Judy Mikovits. Some of the narratives used included that vaccines contain Covid19 virus, masks activate Covid19 virus, and that the Plandemic video was exposing scientific and political elites.

Fake experts are interesting because their credentials lend credibility to outrageous claims. Fake experts use their credential suspend disbelief. Fake experts create an illusion of "another side" of the argument (anti-vaxx, climate change, etc.). It's an effective technique in part because it's a human story. It plays into the narrative of a lone researcher, an outsider, bravely facing off against the scientific and political elites who seek to destroy her and her reputation to maintain the status quo, and it's a story that's cast her as a victim.

Reading

AMITT

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

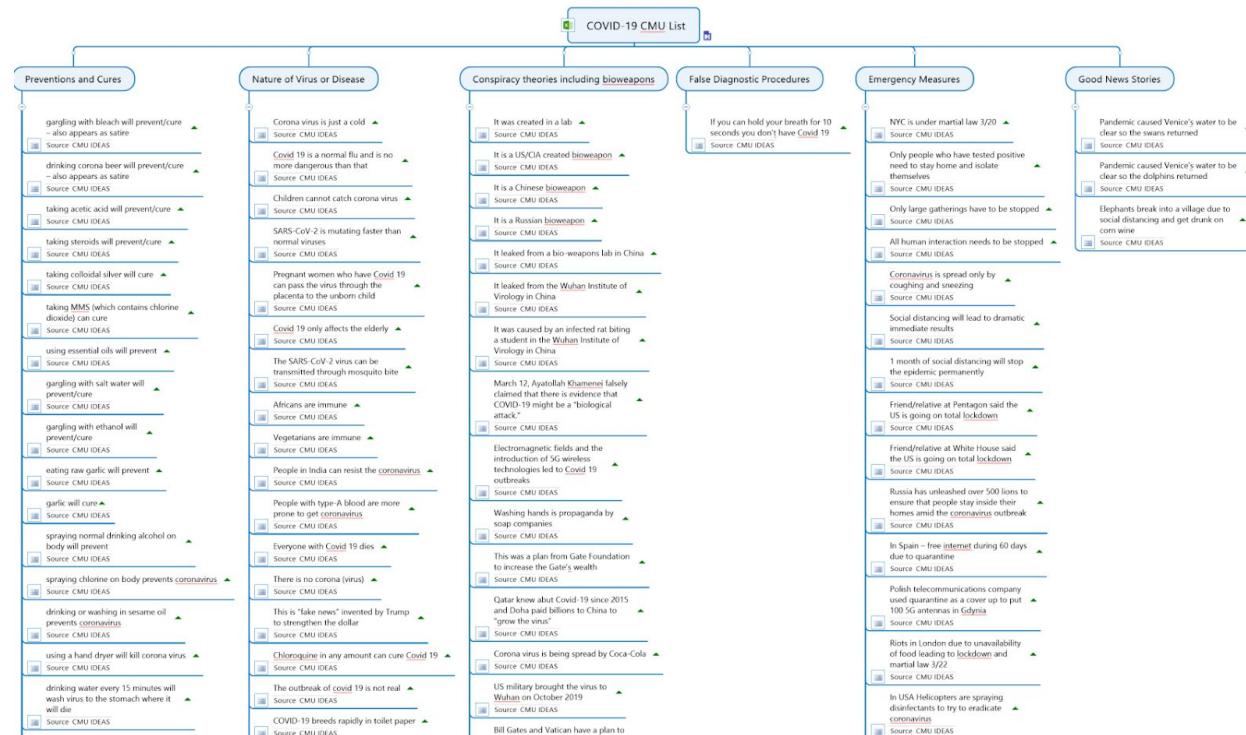
Techniques

- [Russian Election Trolling Becoming Subtler, Tougher To Detect](#)
- [Big Lies and Rotten Herrings: 17 Kremlin Disinformation Techniques You Need to Know Now](#)

Social Media Object Set

STIX gives us artifact object types Observed Data and Indicator, but in MISP we get into more detailed object types like email, url. MISP didn't have a set of objects to cover social media data, so we added a new set with a new object for each new platform type (twitter-post, facebook-group etc). We initially tried using generic objects (social-post, social-group etc), but found these confusing and difficult to work with at speed.

Narrative Models



We know we need to track narratives as they form, combine with other narratives, die away and sometimes reemerge, but we haven't settled yet on a good representation for this. We've tried

mindmaps, and looked at how to match known narratives with the results of things like text-based clustering and anomaly detection.

Chapter 3 Influence techniques

TL;DR Influence Techniques	1
Describing Influence Techniques	1
Strategies	1
Resources	2

TL;DR Influence Techniques

Describing Influence Techniques

The big ideas here are:

- if we have a common description language, we can share information about disinformation incidents in real time.
- if we describe the moves disinformation creators use, especially if they're reused moves, we can mitigate or block those moves

Strategies

The social strategies for mass population influence can be divided into five categories: Distort, Distract, Dismay, Dismiss, and Divide.

- **Distort** the facts. We're not invading Ukraine; we're rescuing/protecting ethnic Russians.
- **Dismiss**: Critics and uncomfortable facts. Make counter-accusations. We've seen this one used often by China. Every time the U.S. accuses China of stealing our intellectual property via illicit hacking, China retorts by first dismissing the accusation and then stating that they're the targets of U.S. hacking.
- **Distract** from the main issue. MH-17 was a tragedy. Why is a commercial airliner flying over a war zone?
- **Divide**: Reduce trust, create confusion, and provoke populations. It's not an accident when two groups at polar opposite ends of the political spectrum "magically" have competing events at the same time and place.
- **Dismay**: Ad-hominem – make personal attacks, insults and accusations. This one is particularly interesting because by even addressing these attacks, you lend them credence. Think about Pizzagate. Making a preposterous claim suggesting that political

elites have a secret sex dungeon full of kids is very hard to defend against without lending the accusations credence.

References:

- Ben Nimmo, 19 May 2015. "Anatomy of an Info-War: How Russia's Propaganda Machine Works, and How to Counter it." StopFake.org 01 December 2018, <https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it/>
- Social graph: Kate Starbird, via <https://twitter.com/katestarbird/status/954805298652332032>

Resources

To implement strategies using the power of social networks, we need accounts with access to social groups, and personas. These types of accounts can be broken into six categories.

- **Bots:** Bulk purchase, mostly amplifiers, little-to-no original content
- **Parody:** Clearly counterfeit account used to satirize or diminish image
- **Spoof:** Counterfeit account which closely copies real account.
- **Camouflage:** False account which mimic community of real accounts
- **Deep Cover:** False account accepted as real for long periods of time
- **Takeover:** Real account in control of adversary

It's easy to get caught up in the technology: hacking accounts, identity theft, botnets, and so on, but it's important to remember that the technology is only one aspect of the integrated problem space.

References:

- <https://medium.com/@timboucher/adversarial-social-media-tactics-e8e9857fede4>

Chapter 4 Looking after yourself

TL;DR Looking after yourself	1
Safety	1
Your Mental Health	2
Basic OpSec for our team	3
Key concepts	3
Process	3
Compartmentalization: Engineering to make mistakes difficult	3
Foundations: Personal Security	4
Foundation: Work environment	4
Work recipes: if this, then that	5
OpSec Appendices	6

TL;DR Looking after yourself

Protect your mental health, protect your online identity, and be aware of the people around you and affected by your work. Repeated exposure to disinformation can wear you down mentally, even if it doesn't look **that** bad. Take breaks, work with other people, eat chocolate, don't use your personal social media accounts to look at disinformation. And lock your shit down.

Safety

If you're going to work on disinformation, you'll need to keep yourself safe. Based on our experience, here are 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 and wear people down. 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, and disgust, as well as to form in-groups and out-groups with hate speech and images that can be difficult to view. This is especially true if the disinformation is about or targeted at 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, a scheduled call from a friend, or the dog pestering you for its walk 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, or 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. 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...

Basic OpSec for our team

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.

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.

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!

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.

Foundations: Personal Security

(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](#)

Foundation: Work environment

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](#).

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.

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.

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 there is no taint or contamination from inside the compartment back to you.

Threat Modeling

From Lorenzo Franceschi-Bicchieri'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."

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](#).

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.

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)

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](#).

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 [\[thatoneprivacysite\]](#) 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 [\[ipleak.net\]](#)(<https://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.

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.

****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.

****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)

****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](<https://voice.google.com/>). You'll obviously need an associated Google account and getting it requires providing a real phone number (major downside).
- Paid VOIP: [Burner](<https://www.burnerapp.com/>), [Hushed](<https://hushed.com/>), [CoverMe](<http://www.coverme.ws/>)

- Burner phones: Lots of different options including [Tracfone](<http://www.tracfone.com/>) where you can get a cheap phone and swap the SIM when needed.

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.

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

Secure Email

End-to-End Encrypted email services:

- [Protonmail](#)
- [Tutanota](#)

Secure Ephemeral Communications:

- [Firefox Send](<https://send.firefox.com/>) 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](<https://cloakmy.org/>): 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.

Social Engineering and Phishing

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

- [Urlscan.io](<https://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'](http://thispersondoesnotexist.com) 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 5 The Disinformation SOC

TL;DR Team Setup	1
The Team	1
Team mission	2
Team needs	2
Specialisations	3
Team Channels	3
Onboarding: coming in to help	5
How Disinformation fits into an Infosec Threat Response team	5
Red Team: Learning how the other half thinks	7

TL;DR Team Setup

- People:
 - create secure space to discuss incidents in
 - ensure team are aware of safety (mental health, opsec)
 - create / get access to training/ mentoring, if needed
- Process:
 - write team mission
 - detail process that's easy to follow under time pressure
 - set up connections to responders/ other teams
- Tech:
 - create safe shared space for notes
 - create / get access to data storage areas, if needed
 - create / get access to analysis tools, if needed
 - create / get access to information sharing tools, if needed

The Team

If you're working on a distributed disinformation defence, chances are you're either setting up a team or part of a team made up of individuals and/or organisations. We've learned a few things about that.

Team mission

You're setting up a distributed disinformation defence team, but you also need to know what its core mission is, so you can check any new requests against it (because scope creep is real, and you will at some point try to handle all the disinformations at once).

As an example, the CTI Disinfo mission is, "We're here to find, analyze, 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."

Team needs

A disinformation response team typically needs:

- People
- Information sharing space / channels
- Incident management process
- Connections to responders and/or other teams
- Supporting technology
- Training / knowledge sharing materials
- An easy way to onboard people / help people find all the materials above (we use the team README for this)

Another way of looking at this is through the process triangle, e.g. we need enough trained people responding fast enough to be able to make a difference to an incident - which includes noticing incidents fast enough, and ways to make those differences happen:

- People
 - Enough people to be able to make a difference to an incident, in the timespan that difference matters (includes noticing incidents in time)
 - Enough connections or levers to make a difference
- Culture
 - Safety processes: mental health and opsec
- Process
 - Understand disinformation, understand threat response
 - Fast, lightweight processes
- Technology
 - Speed - supporting analysis, storage etc
 - Sharing - get data to responders in ways they understand (use whatever works)

We'll cover each of these components, but the most important part of this is people.

Specialisations

Cognitive security needs many different skills - you're likely to find yourself working alongside information security people, but also academics, researchers, technologists, students, and journalists. We've found it useful sometimes to also form specialist teams:

- Leads: 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 the team, arranges resources as needed, coordinates team activities, tracks and logs team activities, and keeps an eye on overall team health (we're also sometimes in a difficult environment, and it helps to have someone watching for stresses).
- Incidents: Runs the disinformation incident response. Prioritizes 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: Makes sure disinfo has all the tech it needs to do its job and keeps that tech running. Builds tools as needed. Finds and guides development talent as needed and appropriate for the disinformation team. Ensures tech builds are documented, repeatable, and maintainable. If appropriate, this role might be accompanied by a research or data lead.
- Process and training: 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_ regular team training, and that the people running it have the resources they need.
- Outreach: 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 and needs overlap with the disinformation team.
- People: 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.

Different teams will have different specialist needs.

Team Channels

One of the features of a team working on disinformation is that it will itself be vulnerable to disinformation and incursion attempts, and will have to design its communications and information sharing channels to accommodate this. In the League, we handled this by running three separate Slack channels for the team:

- Disinformation: an open channel, for people to work on Covid19-related incident tasks, and to learn, engage, and share about Covid-related disinformation and disinformation techniques. Members have full access to open tools, process notes, training materials.
- Triage: a high-trust channel, for people who've been vetted to work on Covid19-related sensitive data and incidents. Members have named access to all data, tools, materials.
- Random: Open channel for anything that doesn't fit into the other channels, e.g. non-Covid19-related resources and observations, general chat about disinformation.

Examples of posts and where they should go include:

- A fun post about disinformation in the gas industry = Random
- Non-medical political disinformation = Random
- Extremist disinformation = Random, unless they're part of a disinfo incident
- An alert about a (non-sensitive) potential new incident = Disinfo
- Articles on health-related disinfo = Disinfo
- Cool tools and ideas = Disinfo (unless sensitive)
- Announcements = Disinfo
- News from other groups = Disinfo
- "How's everyone doing" check-ins = Disinfo and Triage
- An alert about a (sensitive) potential new incident = Triage
- A post about a sensitive incident that we need to keep in triage = Triage

Other channels used in the League include:

- Data channels: useful for streaming supplementary input data that we don't want flooding the main human channels
- User channels: 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: other teams (e.g. darknet) who work alongside us, sometimes on the same artifacts
- Output channels: stream clean outputs from teams

The channel norms we used in the League include:

- Follow the League code of conduct <https://cti-league.com/cti-league/code-of-conduct>
- Don't put disinformation into the disinformation channel without a warning that it's disinformation
- Incidents should be threaded, so please add posts related to incidents in the relevant thread

- Machine safety:
 - Defang your urls (i.e., www\[.\]google\[.\]com)
- Human safety:
 - Keep anything potentially triggering in threads
 - [Content Warnings](#): Live by the rule of “First, do no harm”
 - You can use “CW” to indicate that there’s a content warning
 - Types of content that require content warnings: Violence, Hatred, self harm, etc.

Onboarding: coming in to help

If you're joining a disinformation team, it's on the team leads to make the processes simple to use and tools, notes, training etc easy to find. Some of the helpful things you might need to know include:

- The main work of the disinformation team is incident tracking and response. Live incidents are listed in the incident tracking system, and new ones are flagged in team channels as they're added.
- Everything starts at the team README, which lists where to find tasks and information, the team channels, tech stack, and a cut-down description of the team's incident process.
- Information sources include the BigBook of disinformation response, team-specific explainers, and regular training / knowledge sharing sessions run by the team.
- Triage is a high-trust team, so we vet everyone who joins it. To join triage, fill out the disinformation team survey, so we have enough information to check you're not a bot, sockpuppet or similar.
- When in doubt, ask a team lead: we've created a Slack group, @disinfo-leads, so you can always reach a lead. Otherwise, checking social media to see if a new incident is brewing is a never-ending job.

How Disinformation fits into an Infosec Threat Response team

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

There are disinformation equivalents to these:

1. Neutralize. Disinformation incident response: disinformation triage, takedown, triage and escalation.
2. Clearinghouse. Collate and share incident data, including with organizations focusing on response and counter-campaigns (the “elves” who fight the “trolls”).
3. Prevent. Collate disinformation vulnerabilities and indicators of compromise (IoCs), and supply these to the organizations that we work with.
4. Support. Assess the possibility of direct attacks, and ways to be ready for that. For example, prepare resources that could be used in countering campaigns that target specific facilities, groups and high-profile individuals.

For the neutralization 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), file abuse complaints to registrars, 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.

Red Team: Learning how the other half thinks

If you're defending against a team's actions, it's useful to understand how that team thinks. Whilst it's possible to reach out and speak to disinformation creators, that's not always advisable, so CogSecCollab runs weekly Disinformation Red Team sessions. We've learned a lot from these sessions, often changing the way we work to match insights like "yeah, of course that's why it happens that way".

Some topics you might like to try include:

- running "disinformation as a service"/alternative marketing companies,
- running hostile social media platforms,
- running hybrid disinformation / traditional infosec incidents,
- extending an existing campaign, to predict where it might go next

Chapter 6 The Disinformation Response Network

A Thousand-Point Solution	1
Responders	1
Elves	2
Information Sharing Networks	2
Reaching the networks	3
ISACs / ISAOs	3
MISP Networks	3
Broadcast from your own network	3

A Thousand-Point Solution

"Disinformation isn't a silver bullet problem, it's a thousand-bullet problem" - Pablo Breuer.

Disinformation is a distributed, heterogeneous problem: there are many incident, narrative and artifact creators, across many different channels and communities. The response to this isn't going to be just from a few large organisations: it will need to come from all of society, and be collaborative, heterogeneous, and connected.

Lots of different groups at lots of different scales will need to work together, and we need to connect them, in a way that respects the groups, the subjects of disinformation, and the accounts and groups being investigated. Practically, that means both finding and connecting those groups, and carefully designing around privacy, sharing, and standards.

Responders

Mostly, when people think about cognitive security, they look at platforms, public, and government as responders. But as we catalogued counters, we found many types of people, resources, and groups who could help. A few of these actors include:

- Other infosec people
- Platforms
- Law enforcement
- Governments and government departments
- Communities and Elves

- Influencers
- Media
- Nonprofits
- Academia and Educators
- Industry / corporations
- General public and individuals

Potential responders include the whole of society, including the infosec bodies already linked by the ISAOs and cyber Interpols. You may find yourself sending reports to, or working alongside, many of these people.

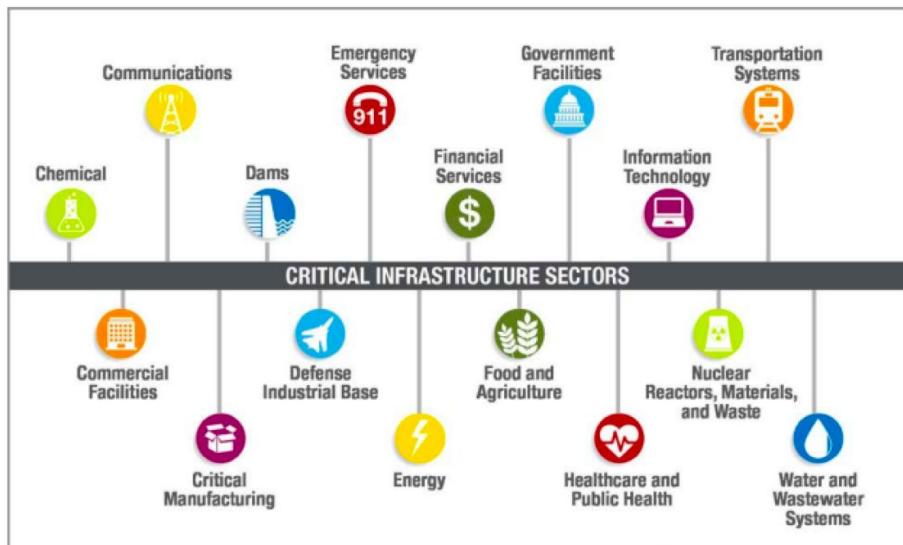
Elves

One group we're focusing on are the Elves. In Eastern Europe, these are volunteer groups who go online to counter Russian troll activities - mostly with tracking, truth and humor. We've been wondering if that could work elsewhere and set out to support it.

References:

[Elves vs Trolls: fighting disinformation in Lithuania](#)

Information Sharing Networks



ISACs (USA)

A group that we support is the information sharing and analysis organizations: the ISAOs, ISACs and cyber Interpols. These organizations already share infosec information for critical sectors in the USA, and we now have one that shares cognitive security information to all the

other ISAOs and ISACs: the CS-ISAO (Cognitive Security ISAO), run by IACI-CERT at the Center for Space Education, NASA/Kennedy Space Center, Florida.

References

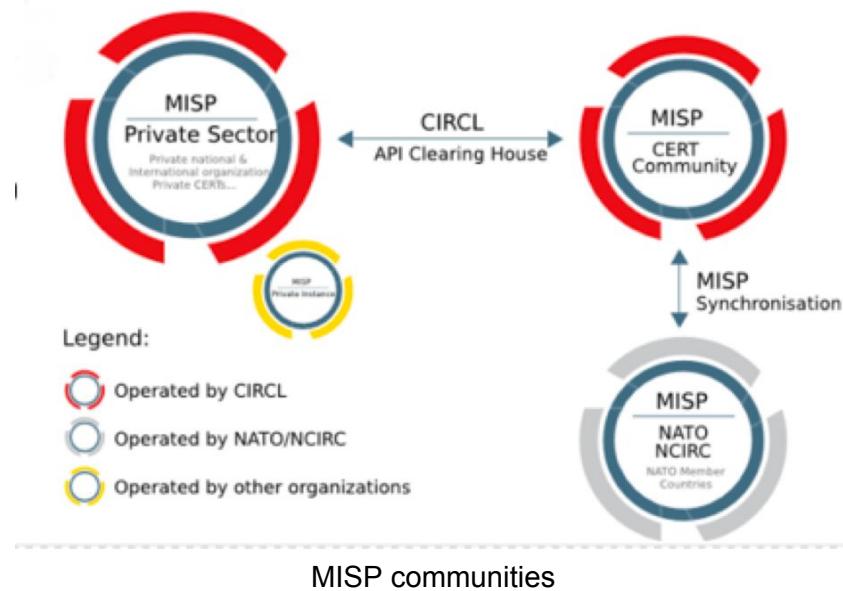
- <https://www.isao.org/information-sharing-groups/enrollment/>
- <https://www.dhs.gov/cisa/information-sharing-and-analysis-organizations-isaos>
- <https://www.dhs.gov/sites/default/files/publications/ISC-PPD-21-Implementation-White-Paper-2015-508.pdf>

Reaching the networks

ISACs / ISAOs

Although your community or organisation can join an ISAO, the easiest way to these quickly is to find someone who's already part of them. Most of them share information using the STIX data standard (see the disinformation models section), many of them use the MISP platform (see the tools section).

MISP Networks



MISP communities of users and organizations run MISP instances that share information about threats and cyber security indicators worldwide. Many are easy to join; it's also possible to stand up your own MISP instance and add it to the network (see Tools chapter). MISP communities include: EU, ISAC, ISAO, CERTs, CSIRTs, NATO, Military, Intelligence, Fortune

500s. Misp connections include API push/pull, email out, and connections to other tools (e.g. Anomali ThreatStream, ThreatConnect, OSQuery).

CogSecCollab runs the MISP disinformation community.

<https://www.misp-project.org/communities/>

Broadcast from your own network

If your community is creating and outputting data, research and other outputs, you could push those out yourself, taking care to respect privacy, not start panic etc etc. In practice, that means sending things like flash alerts through a system that's either part of a larger network (e.g. the MISP networks), or broadcast to subscribers (e.g. email).

Chapter 7 Monitoring

TL;DR Monitoring	1
Monitoring	1
Monitoring Alert Feeds	2
Monitoring Narratives	2
Organising Narratives	3
Identifying new narratives	3

TL;DR Monitoring

Monitor groups, narratives and artifacts (e.g. articles on URLs) in the team's area of interest, so when an incident starts, there's a body of knowledge supporting it.

Monitoring

When an alert comes in, the incident workflow starts (see the next next chapter for details of this). But whilst the sexy incident responses and "took down a huge operation" responses exist, incident response isn't all that a disinformation response team does.

We build out our knowledge bases and communities, write code to speed up our responses, test tools and processes, and work on a lot of the background things that help an incident response run smoother.

Workflows support the top-level mission and goals of the group. Activities to support those goals include incident response, but they also include:

- monitoring for public safety issues that we can report before they become harmful;
- mapping harmful narratives as they emerge;
- monitoring known disinformation feeder channels.
- chasing down disinformation tactics and counters;
- Adding and maintaining supporting disinformation data

Monitoring work includes spreader analysis - looking for infrastructure and accounts that are set up in advance of incidents, including sock puppet accounts "laundered" and left to mature.

Monitoring Alert Feeds

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. It's an indicator that something might be worth investigating and starting an incident response for.

Monitoring Narratives

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. 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.

Organising Narratives

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.

There are a lot of narratives: we've seen hundreds of them in Covid19 alone. We've also seen that these can be grouped into themes; we've used mindmaps to group and organise narratives into hierarchies, making them easier to read and manage. We've also used spreadsheets to share narrative lists.

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 8 Incident Response

TL;DR Incident Response	1
Disinformation Threat Intelligence	1
Incident Response Considerations	2
Response Timescale	2
Level of Engagement	3
Product types	3
Resource Constraints	4
Incident Workflow	4
Make a Go/No-go Decision	4
Triage questions	5
Add incident to logs	6
Alert the Team	6
Create Places to Put Outputs	6
Build Situation Picture	7
Take or Spark Action	7
Help with a disinformation incident	7
Organising an Incident Response	8
Managing an incident response	8
Small-scale system with a Slack and Googledrive	10
OODA loops and intelligence cycles	13

TL;DR Incident Response

Match your response to the time that you can make a difference in, the team and resources you have available, and the effects that you can reasonably achieve. It's okay to say 'no' to response. If you respond, name the incident, decide which questions you're answering, tell the team that you're responding and sort out where response inputs and outputs are going.

Disinformation Threat Intelligence

On a large scale, we see what we do as being part of threat intelligence: the prediction and analysis of recent, current and future threats. To do that, we use:

- **Intelligence analysis**: determine what's going on, who's involved, and what our best guesses are about the current and future situation picture.
- **OSINT**: using public data, determine what we can tell responders, in time for them to act
- **Data science**: support the intelligence analysis and OSINT by finding patterns and information in the data we have available, and help with the "three Vs": the triple problem of data coming in too quickly, at too great a volume, or in too wide a set of formats for the team of people we have available to analyze in the time we have to respond.

Incident Response Considerations

If you're creating a response team, or a new type of response within a team, there are a bunch of axes to think about. How we do this work is evolving rapidly. Another area that also rapidly evolved recently is data science. We've borrowed liberally from data science, intelligence analysis and OSINT practice to help us make sense of this.

Response Timescale

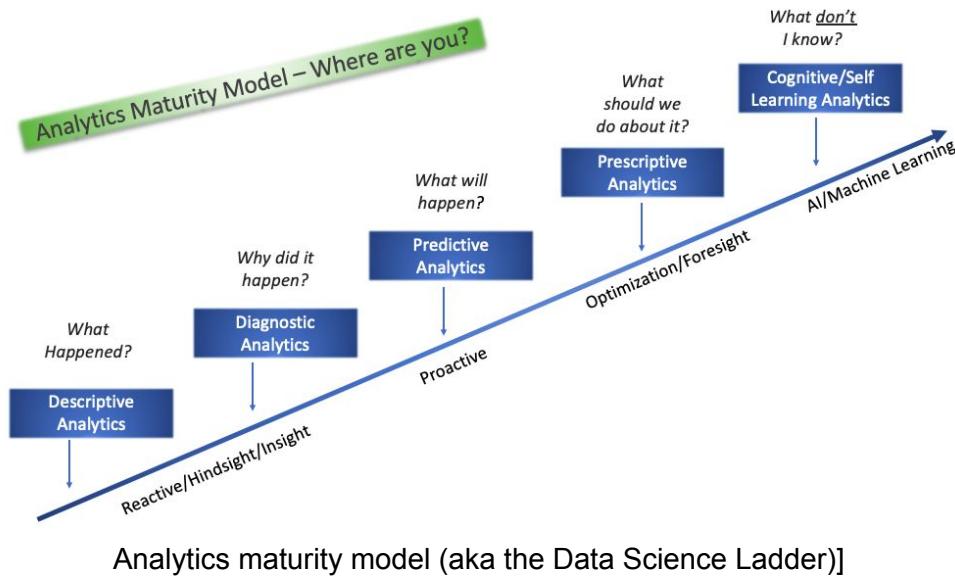
Simply put, how long do you have before your analysis and actions don't make a difference? We've divided both data science and disinformation in terms of response times:

- Strategic - years/months/weeks. Issue focussed (e.g. in-depth investigations, long-form journalism). Good places to look for this type of work include Stanford Internet Observatory, UWashington, Shorenstein Center, Bellingcat, Dfrlab, social media platforms.
- Operational - days/weeks. Project focussed. Examples include data scientists embedded in dev teams, working to answer questions/ build algorithms within software dev cycles. Usually running hypotheses to support things like behaviour-driven and hypothesis-driven development, and lean enterprise (pruning value trees etc). Good places to look include the AI/ML-based disinformation data and tool companies.
- Tactical - hours/days. Incident focussed. Includes data journalists. Good places to look include New York Times, CTI League team, some of MLSEC, some of the crisis-mappers.

Different teams do different things. Some teams are large and fast; others small, slow and deliberate. They're all part of a much-needed response. Interestingly, many of the strategic teams are now becoming more tactical in their response, as their baseline data, tools and connections improve. The TL;DR for this axis is that if you only have hours, all you care about

is stopping the flood; if you have months, you can get into the details of attribution, geopolitics and motives.

Level of Engagement



Data science is sometimes described in terms of the data science ladder (above). It describes the types of work done by teams, from investigating what has happened in the past (e.g. classic statistics), to predicting what might happen next, to suggesting moves and countermoves in an interactive environment. Disinformation response is moving towards the right of this ladder now, hovering somewhere around prescriptive analytics (e.g. groups are starting to both analyse and respond, but not engage in game-like interactions yet). Which part of the ladder will the data part of your response be on?

Product types

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 - Ulndiana
- Academics analysing actors and issues - DFRlab, strategic
- Government agencies / military - strategic
- Commercial interests

What is your team, and the teams around it, really trying to do? What are the objects that are most important to it (artefacts, narratives, actors, intents?) - this will shape what you produce and for whom.

Resource Constraints

Although data science is used to make sense of the large data flows in social media, much disinformation response work is still done by hand and is very similar to classic OSINT.

Just because you want to do it doesn't mean you can (or can yet). Do you have the resources for large-scale data analysis? Is your team recovered enough from its previous deployment(s) to engage in this one?

Incident Workflow

The main workflow in the disinformation team is tracking an incident.

A new 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. The steps that happen next are:

1. Decide whether to start an incident
2. Add incident to logs
3. Alert the team
4. Create places to put incident outputs
5. Build a situation picture
6. Share the situation picture
7. Take or spark action

Make a Go/No-go Decision

An incident needs to be within the team's scope, and large enough to be worth expending team effort on.

Before you start, 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.

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 or 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.

Tracked cases can also include persistent threats - groups, narratives, artefacts etc that are likely to appear in future incidents.

Triage questions

The first questions are about whether to start a response. Since questions usually create new questions, this will also feed into your list of things to prioritise, and/or investigate.

- Is this potentially doing harm?
 - What effects might this have?
 - Is it large? Coordinated? Targetted (to demographics etc)?
- Is this disinformation
 - Is the content false (e.g. misinformation?)
 - Is it e.g. phishing rather than disinformation?
 - Does it include fake groups, fake profiles, fake amplification etc
- Is our team the best one to respond to this?
 - Is this in our area (e.g. CTI = currently working on Covid19 / medically-related?)
 - Is someone else already tracking and responding to this?
 - Do we have the resources to respond?

There are several follow-on questions to these, e.g.

- Is this isn't an incident, is it something that we should handle in a different way? e.g.
 - investigate as part of our monitoring work

- monitor periodically in case it becomes an incident
- If we're not the right team, and nobody else is tracking this, should we:
 - ignore it
 - put out an alert in the hope that another team might pick it up
 - work to find a more suitable team for it

Add incident to logs

Give the incident a memorable name. This helps. Add the incident to whichever system the team is using to track incidents. This gives the rest of the team, who are often on different time schedules, a heads-up that this is an active incident.

Alert the Team

- Put a message in slack, with the artifact 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

At this point, might also send out flash alerts to connected teams too.

Create Places to Put Outputs

- Create a folder in the \[googledrive INCIDENTS folder] for notes and anything that won't fit into the DKAN
- Start adding data to the DKAN

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.

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

Build Situation Picture

Look for related artefacts, accounts, urls, narratives etc

Data we build up in MISP

- Incidents
- Narratives
- Actors
- Urls

Take or Spark Action

- Investigate ways to close down the rumor / repeater sites etc.
- Report on the rumor
 - Add an incident to the MISP instance for this rumor
 - The incident must include some relevant observables such as a Tweet, social media username or URL.
- Write and send notes/reports to the people who can respond
- Close down the rumor and move onto the next one (there's always a next one)

Help with a disinformation incident

- 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)
- 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.

- 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.
- If you spot something significant (new objects tied to the incident etc, new things of interest), update the incident README.

Organising an Incident Response

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 through 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

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.

Small-scale system with a Slack and Googledrive

We've run incident responses armed with nothing more than a Slack group and a Google folder.

With this variant, you still need to track which incidents you're responding to. We used a googlesheet with a row for each incident, giving it a name, a status (live, watching, closed), a start date, an end date (when we closed the response, not when the incident closed), and links to any slack channels and googlefolders we used to collect and store artefacts, and write up reports in.

README_incident_<date>_<incidentname>

Sticky

Artefacts

Actions

After-action notes

Log

<start date> Start

Sticky

Overview

<What is this incident about - what are the risks here, e.g. potential real-world consequences>
<timeframes: are there time bounds on this incident, e.g. events it's gearing towards etc>
<geography/ demographics: any specific targets?>

Artefacts

<summary of main artefacts: could also say "look at MISP" here>

Text and hashtags

- <things you can search for>

Twitter accounts

- <accounts active in this>

Facebook groups and accounts

- <accounts and groups active in this>

Googledoc-based reporting

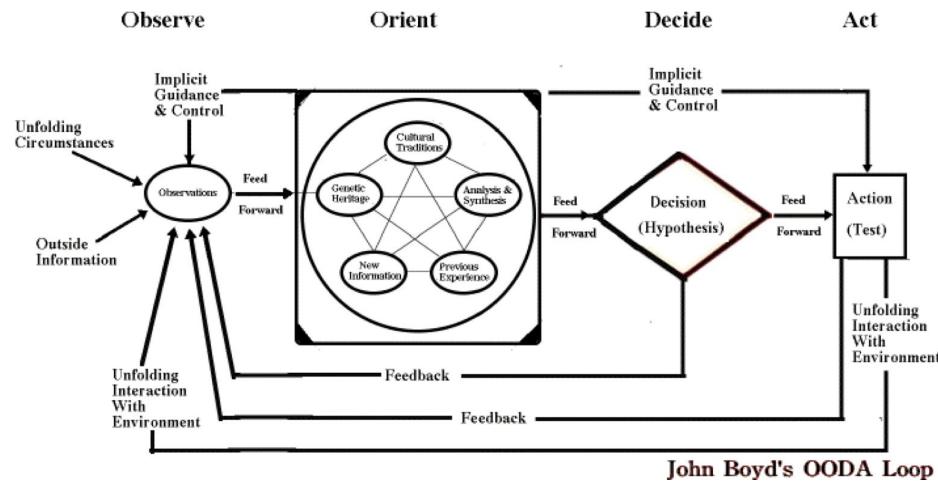
One iteration of these processes used a shared googledrive for each incident, with a folder for images and other artefacts that couldn't be easily copied into a document (videos etc). Within the googledrive was a README file (template above), with sections for artifacts found (because pasting the same image repeatedly is less helpful than giving it a reference number), actions taken, a log of what was found and done on each day of the incident response, and after-action notes when the incident is closed. This is loosely based on some of the shared documents used during crisismapping.

ID	from_object	to_object	Notes
00000001			
00000002			
00000003			
00000004			
00000005			
00000006			
00000007			
00000008			
00000009			
00000010			
00000011			
00000012			
00000013			
00000014			
00000015			
00000016			
00000017			
00000018			
00000019			
00000020			

Googlesheet-based reporting

One flaw of the document-per-incident approach was that it became difficult to share and check incident artifacts. We solved this by created a googlesheet template where we could add artifact URLs, then run code to upload them into other systems for automated analysis.

OODA loops and intelligence cycles



OODA Loop (Boyd)

As we work through incident response, we're going to be mentioning the OODA loop. Boyd's work has been abused for everything from childcare to management theory, but it's still a good way to look at the process from collecting information about the world, through building a "situation picture" that models what you think might be causing those observations, deciding how to act in that situation, then acting and responding to counter-actions in it. First, we'll talk about the "observe" step.

Chapter 9 Detective work: Asking the Right Questions

TL;DR The Right Questions	1
Framing Questions	2
The Five Ws	2
Artifact questions	2
Artefact Tasks	3
Attribution	3
Further Reading	3

TL;DR The Right Questions

Start by listing questions for the team to answer. Most of the time we're doing detective work, looking for evidence that we can send over to another team. Typical questions:

- What artifacts are we starting with? What are they connected to? Have we seen any of these before?
- Who is involved (groups, accounts, etc)?
- What are we trying to do: track artifacts and incident back to origins? Track its spread outwards? Work out where the influence points, places it might be stopped, ways it might be countered or mitigated?
- What are we looking for: just artifacts, or narratives and techniques to add to the incident report?
- Are there other groups already tracking this? Are there related datasets we can pull in and use?
- Who should we be alerting, when, and what do they need?

OSINT, data science and intelligence analysis all have methods that can be useful here.

That's the top-level questions: each artifact will have questions connected to it too, for instance: do we want to map out any networks connected to URLs we find? Are hashtags being amplified etc. These are covered in the next chapter.

Framing Questions

Always start with the questions you want to answer.

- What are we starting with?
 - What's our initial artifact, theme, narrative, lead
- What's our “research question”?
 - What do and don't we care about here?
 - What's more and less important to us (if we have limited resources)?
- What are we trying to produce and for whom?
 - Enough evidence that we can identify who to pass it to, and give them enough to either act or start their own investigation
 - Enough evidence and information to take action ourselves

The Five Ws

In a lot of information gathering and sharing work, we want to know the five Ws: who, what, when, where, why, and how. (https://en.wikipedia.org/wiki/Five_Ws)

Artifact questions

Once we start an incident, our first job is to gather enough information to determine whether we should act, hand this information over to another party, stand down, or not act, but keep a watch on this area. This is usually a mixture of artifact-based activity analysis, network analysis and fact-checking.

- Activity analysis
 - Track artifacts (messages, images, urls, accounts, groups etc), e.g.
 - find artifact origins
 - track how an artifact moves across channels, groups etc
 - find related artifacts
 - Detect AMITT Techniques, e.g.
 - Detect computational amplification
 - Detect, track and analyze narratives
- Network detection
 - find inauthentic website networks (pink slime)
 - find inauthentic account and group networks (including botnets)
- Credibility/ Verification
 - Fact-checking: verify article, image, video etc doesn't contain disinformation.

- Source-checking: verify source (publisher, domain etc) doesn't distribute disinformation.

Fact checking is hard, and usually needs a team of fact-checkers, up-to-date knowledge etc. Source checking isn't the same as article classification, although it does include article classifications. Source-checking is why we label and track URLs. Several groups already publish labelled lists of domains. Fake news creators often run multiple, seemingly unconnected, sites, so finding already-labelled sites in a network can help a lot.

Other related activities include things like looking for signals of inauthentic accounts, inauthentic amplification, and other inauthentic online behaviours, e.g. by looking at patterns of account creation dates for popular messages.

Artefact Tasks

Most alerts start with an artifact: an image, a URL, a piece of text etc. We usually have a lot of starter questions about these too:

- Have we seen this artifact before? Is it related to an artifact we've seen before (e.g. a variant of an earlier artifact).
- Is the artifact's context similar to earlier artifacts (e.g. are the networks and accounts pushing it the same as the ones in earlier incidents?).
- We see a lot of repeat offenders - is this a known scam? Are the actors behind it people already associated with known scams?

Attribution

Attribution - working out who's responsible for a disinformation incident - is hard. You don't have full access to data, and there are incentives for people to obfuscate and hide who they are. At best, attribution is probabilistic, but even a hint can help us assess potential moves, and countermoves.

Further Reading

- Heuer, Structured analytic techniques for intelligence analysis
- Beebe, Pherson, Cases in Intelligence Analysis
- Joint Forces Smartbook, JFODS5

Chapter 10 OSINT: Investigating Artifacts

TL;DR investigating artifacts	1
Introduction	2
Styles and formats	2
Handling Domains (URLs)	3
Chasing a URL	3
Look for ‘similar’ websites	6
Look for new sites	6
Social media references to the site	6
Handling Tweets	6
Chasing a hashtag	7
Chasing botnets	7
Chasing an image	8
Handling Video and Audio	8
Checking video	8
Save an audio file from Facebook Messenger	8
Handling Facebook Groups	9

TL;DR investigating artifacts

First things to do with common artifact types:

- Image: reverse image search; scrape text, urls, hashtags, phone numbers, icons etc and search for each of these
- Tweet: download and analyse twitter search, check images, urls, main accounts from it. Check other social media for related accounts, groups, text
- URL: check registration with icann/whois, use builtwith to look for owner and related sites, use crowdtangle to look for social media mentions. Check site for social media links, owners, sales

- Facebook group: log group name, look for connected groups, look for URLs and other social media accounts, search internet and other social media sites for group name and/or group description text, check page owner / admin

Introduction

Artifacts 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 artifacts that we see most often include:

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

The next layer up includes:

- Narratives
- Domain networks
- Account networks (including botnets)

Many incidents start with an artifact that we investigate, finding connected artifacts, incident objects (e.g. actors and narratives) and techniques in the data available online in social media and other open sources (e.g. typical OSINT inputs). This chapter looks at some of the things we do when we meet different types of artifact.

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?

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, we explicitly declare the timezone or use UTC.

Handling Domains (URLs)

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](<https://gist.github.com/stevenswafford/393c6ec7b5375d5e8cdc>)! (h/t to Roger)

Using Google Dorks to Check Primary sources (from Henk van Ess's [Finding patient zero](<https://datajournalism.com/read/handbook/verification-3/investigating-actors-content/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 its 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](<https://whois.domaintools.com/>) 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](<https://ahrefs.com/backlink-checker>) or [smallseotools](<https://smallseotools.com/backlink-checker/>) 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"](<http://overcognition.com/2020/01/22/data-safari-rough-notes-pink-slime-network/>)

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 generating typosquats include dnstwist

- Near-duplicates [SnaPy] (<https://github.com/justinnbt/SnaPy>)
- [typosquatting] (<https://github.com/topics/typosquatting>)

Feed of domains that were created each day: [whois
newly-registered-domains] (<https://whoisds.com/newly-registered-domains>)

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

Look for new sites

Github code check_new_registrations.ipynb searches for strings of interest in newly-registered domains (from [whois
newly-registered-domains] (<https://whoisds.com/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.

Social media references to the site

Crowdtangle [chrome extension] (<https://apps.crowdtangle.com/chrome-extension>) will give you a list of references to a site you're looking at, on Facebook, Twitter, Instagram and Reddit.

Handling Tweets

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.

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.

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])(https://en.wikipedia.org/wiki/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](#).

Handling Video and Audio

Checking video

[InVID__EU] (https://twitter.com/lnVID_EU)

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"

Handling 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 hand 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 11 Data Science: Investigating Big Data

TL;DR Collect Raw Data	1
Data science	2
Disinformation Modelling at Scale	2
Tactical data science	3
Where and how to look for examples	3
This is where the data science comes in...	3
Where does Disinformation Data Come From?	3
Types of data	4
Data inputs: Alerts and Canaries	4
Data sources: disinformation data streams	5
Narratives	5
Data	5
Counter-disinformation feeds	5
General disinformation datasets	6
Collecting your own data using tools	6
Twitter data	6
Facebook data	7
Reddit	7
Multi-platform tools	7
Storing datasets	7
Disinformation Data Science Courses	7

TL;DR Collect Raw Data

Identify available relevant datasets

- Existing collections (NB their collection isn't your collection: be aware of biases, data gaps etc)
- Social media feeds: searches, APIs, and spreadsheets

Extract data

- Get data onto your machine / into the team's datastores
- In a format you can use: pdfs aren't usable data, and a spreadsheet isn't a MISP object...
- check that your data is clean

Data science



image from

<https://medium.com/the-mission/deconstructing-data-science-breaking-the-complex-craft-into-its-simplest-parts-15b15420df21>

Data science is a process. There are many versions of this process, but it basically goes from identifying and asking a set of business questions, to attempting to answer them by finding and cleaning datasets, building models based on them, and using those models to understand part of the world and/or predict what might happen next in it, then explain that the people who need to make decisions based on your findings.

The start and end of data science is all about people: we need to think in terms of end-users, questions and problems.

Disinformation Modelling at Scale

There's a lot of academic work on modelling disinformation at scale. Some of the models used include:

- automated fact checking
- cascade and time-based models
- social network analysis \((Pablo is keen on scale-free networks for this)\)

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.

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”.

This is where the data science comes in...

Where does Disinformation Data Come From?

The cynical amongst us would say that we're drowning in disinformation data. Mainstream news has many stories about disinformation incidents and takedowns, political groups are quick to decry "fake news", and almost everyone working on disinformation has a favourite fake cure or conspiracy theory.

In practice, if we're looking for disinformation in our specific areas of interest \e.g. the CTI League currently works on Covid19 related disinformation\ in time to make a difference to its effects, we need to do some groundwork and build out connections, information feeds and catalogues of good places to look.

Types of data

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.

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 (\(canary, as in "canary in a coal mine"\)).

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.

Narratives

- EuVsDisinfo database <https://euvdisinfo.eu/disinformation-cases/>
- [Ryerson Claimwatch dashboard](#)
- [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)
- [Indiana University OSOME Decahose](#)

Counter-disinformation feeds

- Snopes: (<https://www.snopes.com/>)

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](#))

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 \(\text{e.g. if there’s a WHO page on a known scam, that might be enough evidence to ask for takedowns etc}\). But most of the time, you’ll have to go collect your own data from across social media, and sometimes beyond \(\text{e.g. for paper flyers, we asked people if they’d seen them in their neighbourhoods too}\).

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

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 gephigraph input data so you can create user-user etc graphs (see the gephigraph 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.

Useful references on collecting twitter data include

<https://firstdraftnews.org/latest/how-to-investigate-health-misinformation-and-anything-else-using-twitters-api/>

Facebook data

The Facebook API is horrible. Most everyone tracking social media uses a third party like [CrowdTangle](#) or scrapes for the data they want. The Crowdtangle chrome extension is available free to anyone, but the full Crowdtangle tool isn't: it's available to news organisations, some academics, and pilot programs, so it's worth checking to see if your [team is eligible](#) or has someone with access on it.

Reddit

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

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](#) and an ELK stack <https://www.elastic.co/>.

Disinformation Data Science Courses

The CogSecCollab / CTI League 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.

Chapter 12 Machine Learning: Pattern Hunting

Notes	1
Introduction	1
AI Overview	2
Algorithms: Text Analysis	2
Narrative detection and analysis	2
Algorithms: Network Analysis	2
Image, Video and Audio Analysis Needs	3
Test datasets	3
Further reading	3

Notes

Data science nice, but big volume, patterns not obvious: use machines. AI good on other side, e.g. deepfakes; also for detection etcetc - Pablo's talk on AI/ML and disinfo.

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.

In the League, we also do many things by hand, but are working on ways to speed them up, and automating the parts that make sense (this will probably never be a fully-automated activity, but we can support and get a lot of load off disinformation analysts) as we validate process.

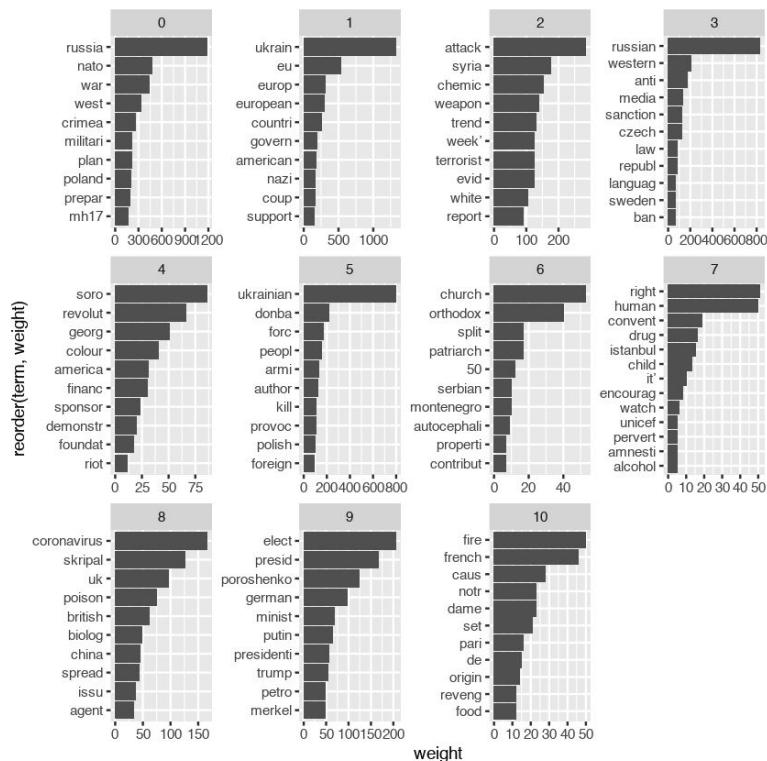
This chapter is about tackling the three Vs (volume, velocity, variety) with machine learning, automating the work that's too large, coming in too quickly or across too many channels for the team to concentrate on them all. There is no magic "plug in this algorithm and disinfo will go away" system: the big idea here is that the humans and algorithms can work together - e.g. this is augmented intelligence that lets the people focus on what they're good at, not a replacement for people.

One of the reasons we shift work from humans to algorithms is make sifting through the data and highlighting potential patterns in it more efficient; another is that disinformation artefacts (images, text etc) are often difficult to handle, and we want to reduce exposure to them as much as possible. If we can cluster artifacts so that instead of looking at 100 near-identical images, a human can view and classify one copy of that image, the reduction of stress on the humans is worth is.

AI Overview

<Fixit: drop in Pablo's AI talk>

Algorithms: Text Analysis



co-occurrence graph for text clusters in EuVsDisinfo (Gabe)

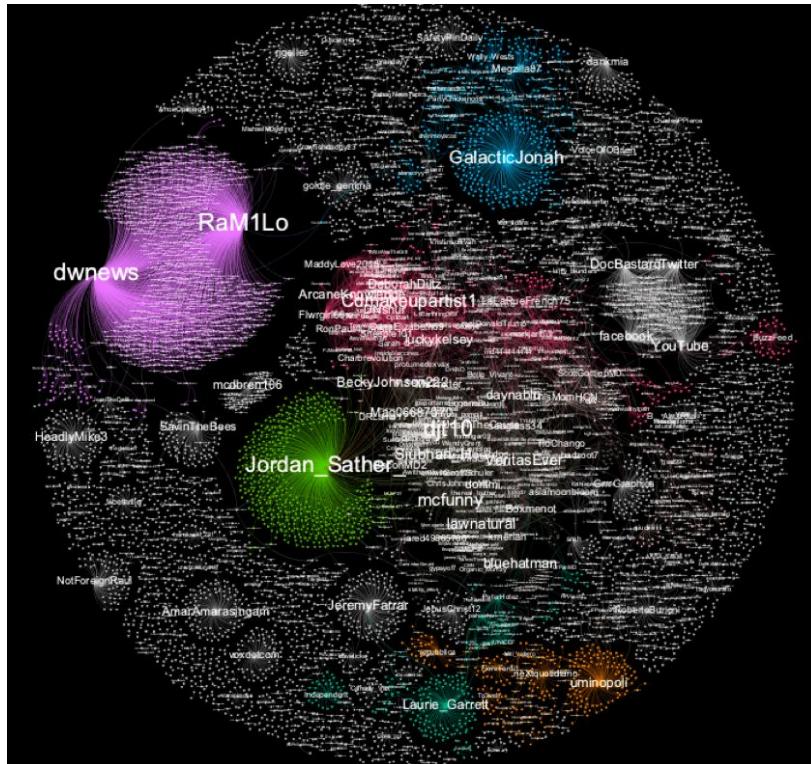
Text-based algorithm needs include:

- Find themes
- Classify to narratives
- Cluster text to narratives
- Search for similar text/narratives

Narrative detection and analysis

Topic modelling.

Algorithms: Network Analysis



Gephi visualisation around a Twitter hashtag

Network algorithm needs include:

- Finding super-spreaders
- Finding rumor origins
- Uncover new artefacts
- Track movement over time

Image, Video and Audio Analysis Needs

Image/Audio algorithm needs include:

- Cluster images

- Search for similar images
- Detect shallowfakes

Whilst we've seen deepfakes being used for things like fake profile pictures, most of our image/video etc needs have been more mundane: searching for reused and/or mistagged images, finding images that have been crudely doctored (shallowfakes), and clustering sets of near-identical images to make them faster to sift through with less exposure of potentially-harmful material to the people checking through them.

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](#)

Further reading

- https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU%282019%29624278_EN.pdf
- [Attention is All They Need: Combatting Social Media Information Operations With Neural Language Models](#) - Fireeye on text generation and detection
- Looking for more: Google search “disinformation ‘data science’” - lot of posing. Found data science sites’ articles on disinformation. Github searches for “[disinformation](#)” and “[misinformation](#)” found hundreds of repos. Searching student sites like Towards Data Science for “[fake news](#)”, disinformation, misinformation can be useful because they’re scanning recent work.
- <https://www.cnn.com/2019/05/23/politics/doctored-video-pelosi/index.html>
- <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>

Chapter 13 Situation Pictures

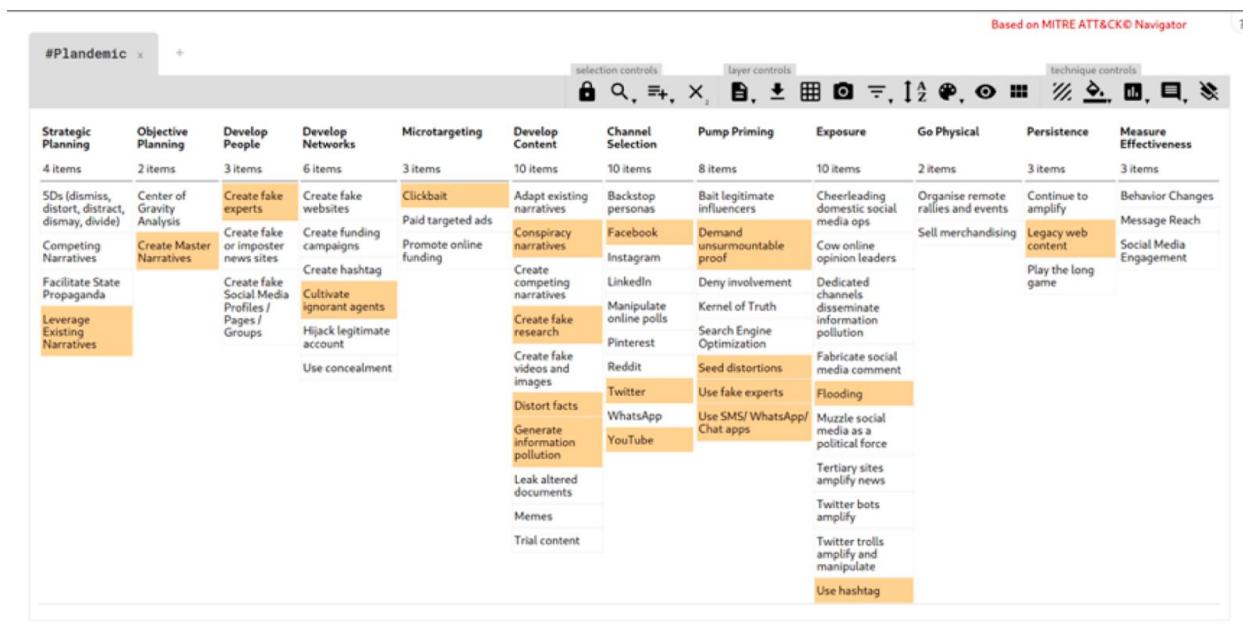
Using TTPs

1

You have questions, artifacts, a community that you're sharing with. You still need to build a picture of what is happening - the situation around those artifacts, that most likely created them - and share that with other people.

Sensemaking includes looking at what we've collected, to work out what's happening and might happen across the whole incident. One way we do that is by analyzing the connections between incident objects. The CTI League uses MISP to help with that; other teams use tools like Maltego.

Using TTPs



TPP framework for Plandemic, 2020

We've talked about the AMITT Framework before. It's how we break an incident into techniques that we can analyze and counter.

We tick the AMITT boxes whilst we're gathering data (e.g. during the observation part of an OODA loop). During Orient, we look at this diagram to work out what's happening, how we might respond, and, if we catch an incident early, which downstream techniques might be used in that incident too.

The example here is Plandemic - a debunked conspiracy theory video which makes some false claims about the nature of COVID-19. We mapped it in AMITT to help us understand what capabilities the actor has and potentially how they're resourced.

Chapter 14 Sharing Information

Making analysis outputs usable	1
Outputs	1
Reports	1
MISP events	1
Output components: visualisations	2
Understanding time series	2
Understanding relative sizes	4
Understanding connections	6
Others	7

Making analysis outputs usable

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.

There's no point building things without thinking about the end users, so 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.

Outputs

Reports

What we get out of Orient is an incident report, containing a summary, narratives, techniques, artifacts and objects.

MISP events

We get a misp event that we can share with other groups either directly or by email, via their threat intelligence tools etc.

We added a few other things to MISP for this.

- Object types for common social media platforms, and code to load these into MISP using single-line commands in Slack, because speed is everything in a tactical response.
- New relationship types, to make the graphs that users can traverse in MISP richer.
- Taxonomies to cover things like types of threat actor.

Output components: visualisations

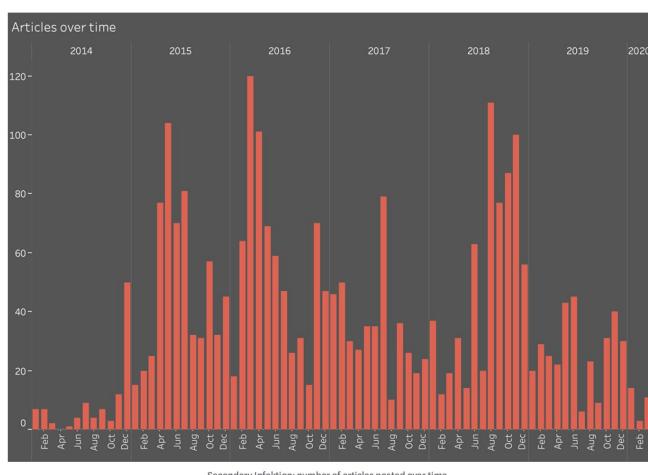
Eyeballing the data, looking at statistics, and examining machine learning outputs are good, but part of getting to know data, and explaining it to other people is being able to look at it visually. There's a lot of work on data visualisation (read "Storytelling with Data" to see it done well), so this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

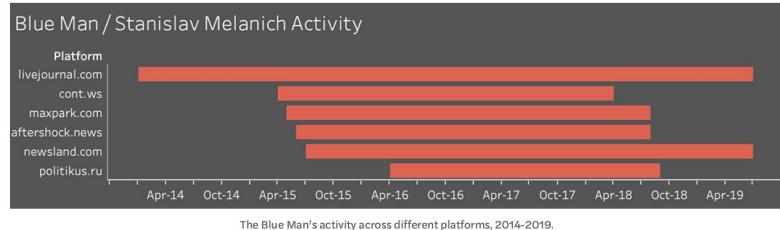
Understanding time series

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](https://matplotlib.org/3.2.1/api/_as_gen/matplotlib.pyplot.bar.html#))



(Sekondary Infektion report, 2020)

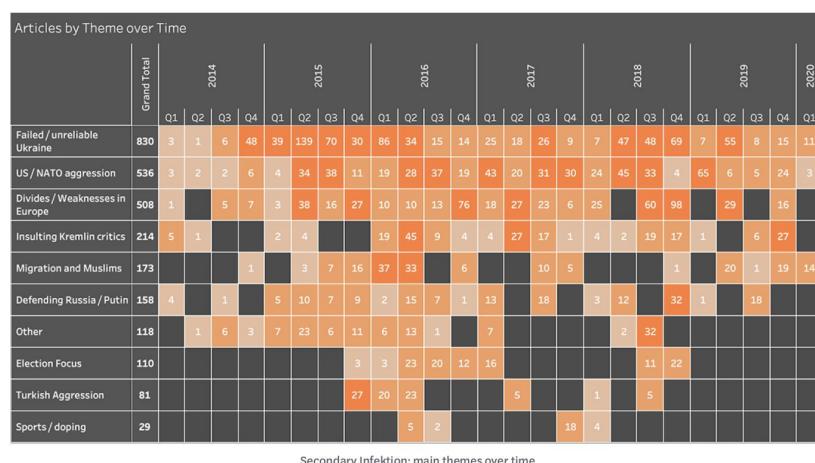
Bar graphs and line plots can be used for showing a range of entities over time.



![(Sekondary Infektion report, 2020)]

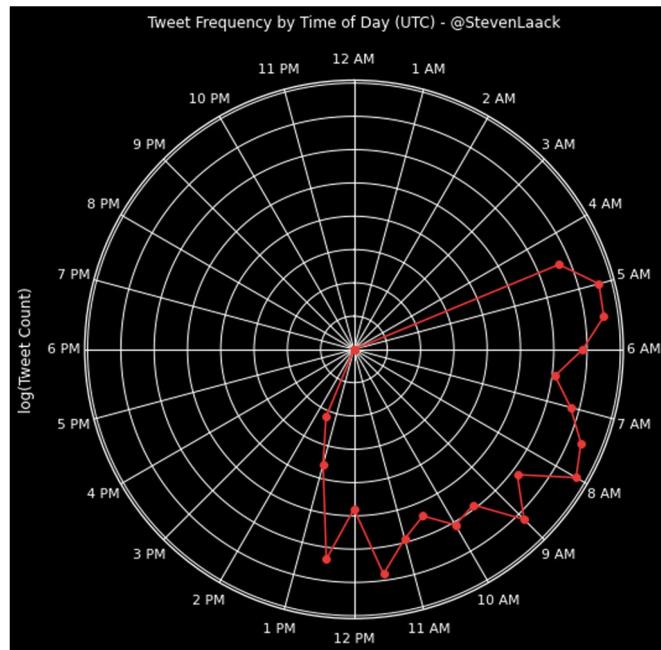
If the value range is too large to show easily (e.g. there's a mix of very small and very large values that you can't easily plot on one axis), heatmaps might be more appropriate.

[<https://python-graph-gallery.com/91-customize-seaborn-heatmap/>] (<https://python-graph-gallery.com/91-customize-seaborn-heatmap/>)



![(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.

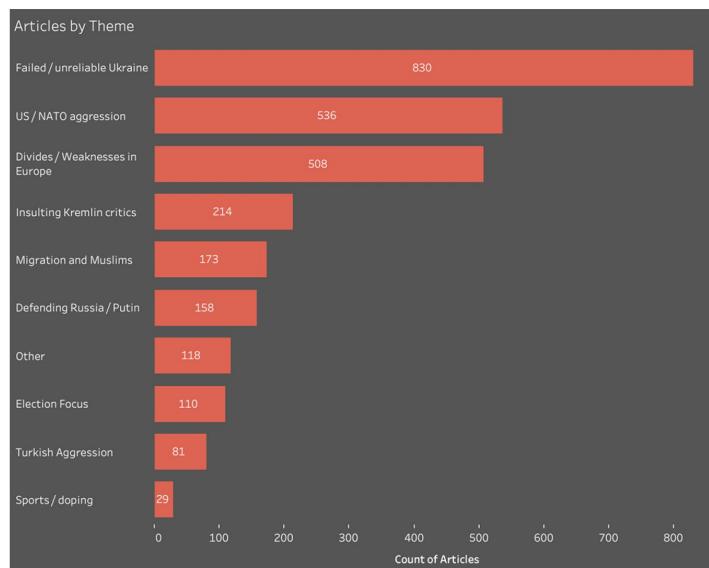


![(Sekondary Infektion report, 2020)]

Understanding relative sizes

Bargraphs can do this too.

[<http://python-graph-gallery.com/barplot/>](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.

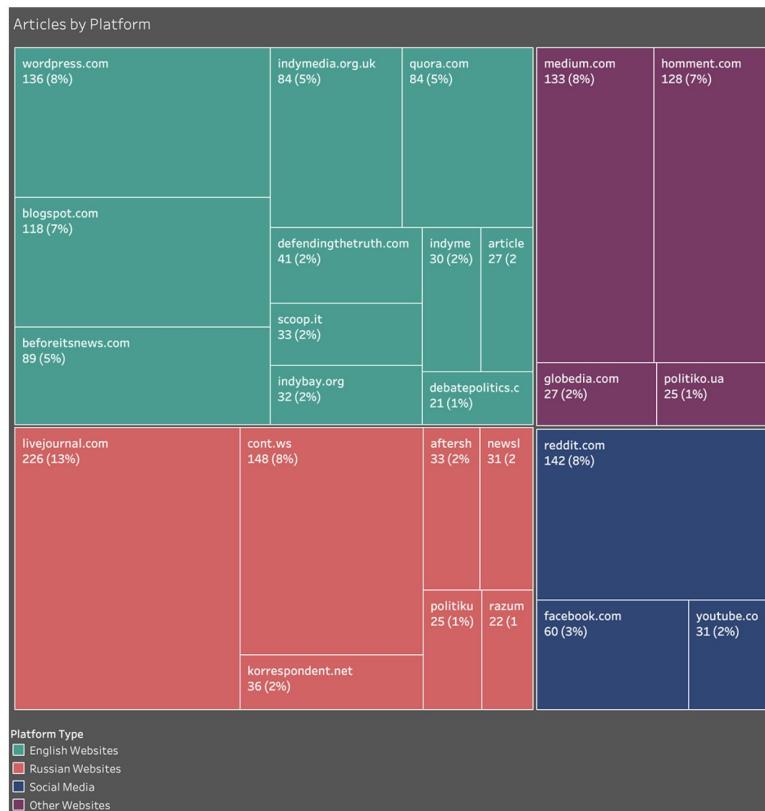
[<https://python-graph-gallery.com/200-basic-treemap-with-python/>] (<https://python-graph-gallery.com/200-basic-treemap-with-python/>)



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

Sekondary Infektion report, 2020

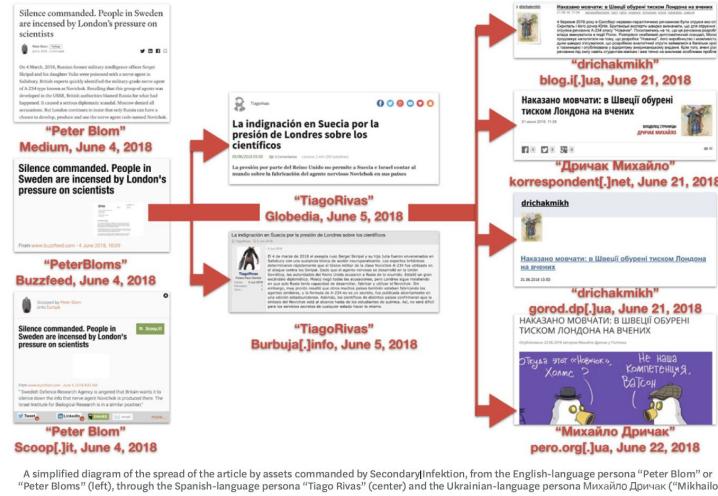
Colours are an extra, useful, dimension on most plots.



Sekondary Infektion report, 2020

Understanding connections

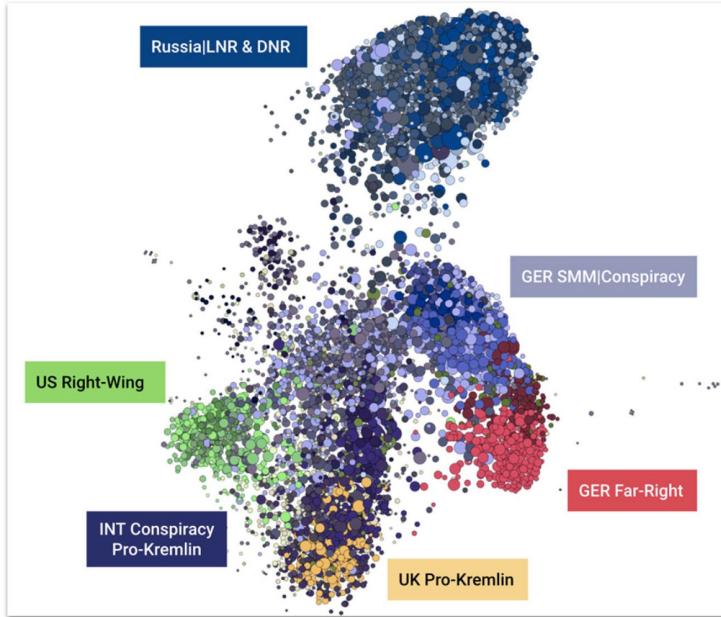
Really simple graphics \think powerpoint\ can help explain the connections between objects.



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 15 Taking Action

Introduction	1
Action	2
Countermeasures	3
Effects-based countermeasures	4
Example: CTI League	4
Doctrine-based Countermeasures	5
Example	5
Playbooks	6
Practical Countermeasures: External Actions	7
Government	7
Practical Countermeasures: Reporting	7
Reporting inside the League	7
Reporting to law enforcement from the League	7
Reporting to platforms	7
Reporting websites	8
Practical Countermeasures: Direct Action	9
Further Reading	9

Introduction

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

- Not acting
- 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.

Not acting. This is always an option: we should always ask if we should act, and if we want to act (and if not, what are the ethical ways we have to discharge responsibilities like having the datasets that we have).

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 law enforcement. Escalating to law enforcement is appropriate especially if there is risk of physical harm, but use this route wisely.

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.
- Takedown requests will need a reason, the easiest of which is violations of platform terms of service. This is about pointing platforms at incidents, artifacts and behaviors they might not have detected already; it's also about countering disinformation incidents: we are not censors, and should always view data in terms of "what is happening" rather than "I disagree with this post".

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/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> - we’re expecting to uncover a bunch more as more infosec people do disinformation.

Countermeasures

"Countermeasures are that form of military science that, by the employment of devices and/or techniques, is designed to impair the operational effectiveness of enemy activity."

Countermeasures can be active or passive and can be deployed preemptively or reactively." - JP 3-13.1 , Information Operations - Joint Chiefs of Staff

The MisinfosecWG collected and designed countermeasures against AMITT tactics and techniques in 2019. We did four main things to get our list of 200-odd counters.

- Existing: We looked for existing countermeasures, in incidents, literature and examples.
- Tactic-based and Technique-based: We ran a workshop, where we used the AMITT framework to create counters to both the tactic stages, and to specific techniques within those stages. This centred around a courses of action matrix; basically we gridded out response types and tactic stages, and asked people to post ideas into each grid square. (our team was known for using all the postits in the building).
- Doctrine-based: And we looked at influence operations as resource-limited games, and described the counters that we could use to deplete or exhaust disinformation resources.

We found quite a few existing counters, beyond the obvious “take down the botnets” and “educate people” ones. Examples included:

- The Macron election team’s email honeypots,
- US Cybercommand blocking the Internet Research Agency’s internet access during the 2018 midterm elections.

Effects-based countermeasures

Our courses of action matrix used a subset of the effects listed in JP3.0:

- Detect: find them. Discover or discern the existence, presence, or fact of an intrusion into information systems.
- Deny: stop them getting in. Prevent the adversary from accessing and using critical information, systems, and services.
- Disrupt: interrupt them. Break or interrupt the flow of information.
- Degrade: slow them down. Reduce the effectiveness or efficiency of adversary command and control or communications systems, and information collection efforts or means.
- Deceive: divert them. Cause a person to believe what is not true. military deception seeks to mislead adversary decision makers by manipulating their perception of reality.
- Destroy: damage them. Damage a system or entity so badly that it cannot perform any function or be restored to a usable condition without being entirely rebuilt.
- Deter: discourage them.

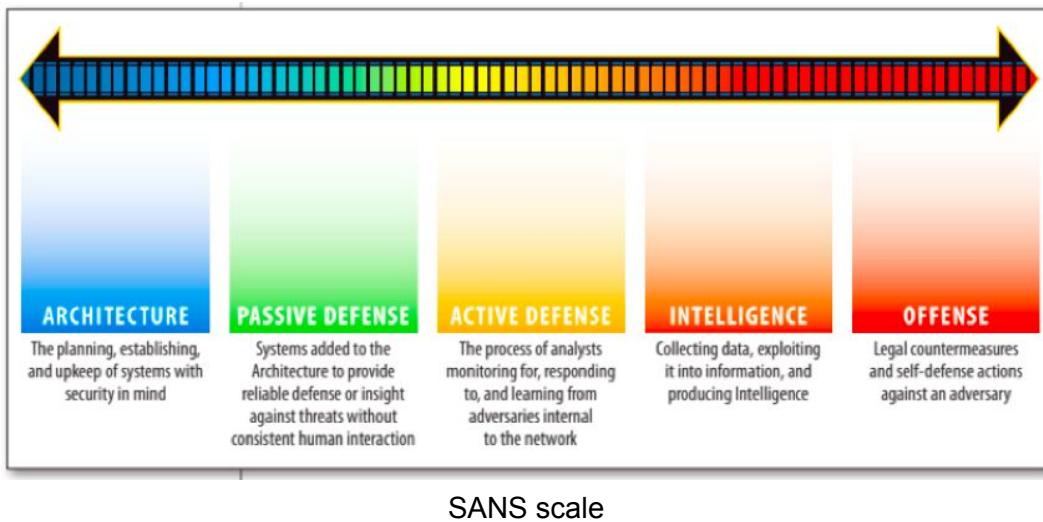
We included Detect because that's what everyone was doing - looking, not reacting, and we wanted them to get that out of their systems. We added Deter to the list as a potentially useful category too.

AMITT metatechnique courses of action

	ALL	D1 detect	D2 - Deny	D2 Deny	D2 deny	D3 Disrupt	D3 disrupt	D4 Degrade	D5 Deceive	D6 Destroy
cleaning	0	0	0	2	0	1	0	1	0	0
countermessaging	0	0	0	3	0	7	1	4	0	1
data pollution	0	0	0	0	0	1	0	4	1	0
daylight	0	1	0	5	1	8	0	2	0	1
dilution	0	0	0	0	0	5	0	1	0	0
diversion	0	0	0	2	0	10	0	2	3	0
friction	0	0	1	12	0	6	1	6	0	0
metatechnique	4	0	0	3	0	6	0	0	0	0
reduce resources	0	0	0	2	0	1	0	1	0	0
removal	0	0	0	14	1	4	0	0	0	0
resilience	0	0	0	8	2	7	0	7	0	0
scoring	0	0	0	7	0	0	0	0	0	0
targeting	0	0	0	1	0	6	0	3	0	0
verification	0	0	0	2	0	1	0	0	0	0
TOTALS	4	1	1	61	4	63	2	31	4	2

AMITT countermeasures, listed by metatechique

The result was a set of mitigations and countermeasures that we labelled by the tactic, technique, agents who could carry them out, and a metacategory that we used to sort through looking for duplicates.



We also thought about these effects in terms of the SANS scale for responses - e.g. whether they were architectural changes to the underlying ecosystem, defence, intelligence gathering or offence.

The AM!TT effects-based countermeasures work can be found in <https://github.com/cogsec-collaborative/amitt>. We're still getting the countermeasures into usable order: you can track our progress in this repository, through the grids that we're using to think about the types of blue-team actions that can be \and are\ used against disinformation.

Example: CTI League

The CTI League uses effects-based counters: reporting to law enforcement, platforms, and registrars, with the CogSecCollab helping to set up the RealityTeam counternarratives group to help counter rapidly-evolving narratives.

Doctrine-based Countermeasures

"A disinformation campaign is made up of resources and infrastructure and operates over time, with time as a universal scarcity." - The Grugg

We need some way to make counters work well together. We have a set of counters but we need a way to understand how and when to use them. SJ, Gruggq, Pablo started a conversation about resources, infrastructure, and that time is scarce. Operations are rooted in the real world

and the critical elements required in the real world limit these actors. Actors need money, people, organization, knowledge and capabilities and the time to make things happen.

When we talk about a counter for a technique we just don't want any viable counter. We want the counter that's most appropriate to our achieve our objectives within the current environment. Maybe that's to destroy some capability, or cost the adversary money, waste their time in some way that is impactful to that adversary based on that actor's state in the real-world.

So what are the critical elements we want to affect? We like to get it RITE:

- Resources: material things, money, messages, audience
- Infrastructure: media/medium, administration, observation
- Time: speed, capacity, "mythical person month"
- Execution: actors, capabilities, strategies

Now we can build a course of action matrix based on critical elements and combine it with the course of action matrix for Amitt techniques. When action is taken on critical elements, a capability's capacity to influence a target audience is affected in some strategically significant way. Effects on disinformation capabilities can now be grouped into three main classes;

- those which exhaust the capability's resource dependencies;
- those which decrease a capability cost-effectiveness; and
- those which exhaust the adversary's capacity to use a capability in a timely manner.

Example

DOCTRINE-BASED COUNTERMEASURES
IRA IN GHANA: DOUBLE DECEIT

- Resources
 - Staff ~16
 - Audience ~338k
 - Mobile Devices
- Infrastructure
 - NGO
 - Operator Content Pool
 - Twitter Analytics
- Execution
 - T0007, T0010, T0015, T0055, T0013
 - T0014, T0018, T0021, T0030, T0039
 - T0042, T0053
- Time
 - Direct Engagement
 - No Automation + Bots
 - 'Audience Building' Phase

Double Deceit incident components

That's the theory, can we apply that in practice? Turns out we can. Double Deceit is interesting to a critical elements based doctrine of counters for several reasons: Small operation: 16 people around a table; essentially a bunch of kids with phones. complicated situation for a defender - can't target an NGO without understanding who they are and what they're doing; they might be legitimate, but appear adversarial in nature.

So what can we do the next time we detect this pattern? Banning the accounts and alerting them just teaches them about our own capabilities. Ideally we want to make them ineffective but allow them to operate. Target critical elements the adversary is weak in and turn this IO into a resource sink for our adversary.

The critical elements of this operation can guide our application of counters. For example,

- this a highly time constrained operation. No bots. Direct engagement. Bottlenecks in how we could expect them to react to our counters. We can slow them down and make them ineffective.
- Another vulnerability is that they relied on Twitter Analytics and understanding retweets, messages. It appears that their KPIs are all contingent on social media platforms giving them accurate results.

Playbooks

Threat Intelligence playbooks work towards a goal. We have something we want to protect, achieve, deny etc. Playbooks can build complex responses to disinformation events. Can tell us how to respond to an adversary given a set on conditions and objectives.

In double deceit, we saw time was one vulnerability. Double Deceit was also vulnerable in how it collected and used analytics. If we want to disrupt that, we could do something like a fake engagement playbook - a set of actions to disrupt that.

Title	RP_0003_fake_engagement
Description	Response playbook for effects on social media engagement analytics.
AM!TT Tactic	<ul style="list-style-type: none"> • TA03: Develop People • TA06: Develop Content
Tags	<ul style="list-style-type: none"> • amitt.T0007 • amitt.T0020 • amitt.T0021
Severity	Low
TLP	GREEN
PAP	WHITE
Author	@VV_X_7
Creation Date	17.03.2020
Detect	<ul style="list-style-type: none"> • C_00223_interview_ignorant_agents
Disrupt	<ul style="list-style-type: none"> • C_00135_deplatform_online_community
Degrade	<ul style="list-style-type: none"> • C_00103_engage_with_nlp_bot • C_00104_engage_with_elves
Deceive	<ul style="list-style-type: none"> • C_00220_fake_engagement_system_amplify_impression • C_00221_fake_engagement_system_amplify_engagement • C_00222_fake_engagement_system_use_fake_persona

Workflow

1. Execute Response Actions step by step.

Example playbook \fake engagement\

A fake engagement playbook looks something like this. We list the relevant adversary capabilities and the set of appropriate counters to achieve our objective. What we're doing here is building a playbook that lists the possible effects we can have on the adversary capabilities and which counters we need to use to achieve that effect.

Our next step is integration of critical elements to guide decisions to deny, degrade, etc., by the type of resources required for the capability. For example, we could degrade adversary use of analytics by targeting time or resource sensitive requirements for those analytics. At which point we start getting into game theory.

Practical Countermeasures: External Actions

We can't always act ourselves, but we often know an organisation or group that can, given direction on where to look, and/or reliably gathered evidence. Groups like CS-ISAO also need ways to share disinformation information quickly between organisations.

Government

Table 1: Counter-disinformation strategies used by the three institutions in this paper, and their effectiveness and legitimacy in a democratic society.

Strategy	Used by	Effectiveness	Legitimacy
Refutation	EU Stratcom Facebook via fact-checkers	Works if consistent, but not all disinfo is about facts.	Generally legitimate to speak the truth, though people will disagree on what truth is.
Expose inauthenticity	EU Stratcom Facebook	Discredits the source, provides justification for further measures.	Content-neutrality is appealing. Important to preserve legitimate anonymity.
Alternative narratives	EU Stratcom China	Helps displace disinfo, inoculates against it if seen first.	Can itself be disinfo or distraction.
Algorithmic filter manipulation	Facebook China via 50c party	Media algorithms have huge effect on information exposure.	Platforms may abuse this power, users may game it.
Speech laws	Facebook enforces such laws China	Can be effective at targeting narrow categories of speech.	Broad laws against untruth are draconian.
Censorship	China	Effective when centralized media control is possible.	Generally conflicts with free speech.

Jonathan Stray, “Institutional Counter-disinformation Strategies in a Networked Democracy

Jonathan Stray's survey on government response is useful here.

Practical Countermeasures: Reporting

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.

Reporting to law enforcement from the League

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

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](#)
- [Unreliable News](#) repo , [Cred Score](#) (hypothesis), [Fact-check Feed](#) (articles by US fact-checkers, 2016–present), [Fact Checkers](#) tool, [News Netrics](#) media site performance metrics.
- <https://iffy.news/fact-check-search/>

Practical Countermeasures: Direct Action

Some groups (e.g. the kPop stans) have taken direct actions against disinformation, including flooding hashtags with external material (band photos etc).

When you act in a disinformation space, you're acting in an environment, with a lot of other humans and machines in. And what you can end up in is a multiplayer game, where you're each acting in response to each other, and playing off against each others' resources. Be aware of this if you choose this route.

Further Reading

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

Chapter 16 Tools

Response Toolkit	1
Incident Tracking Tools	2
HIVE	2
Incident Sharing Tools	2
MITRE ATT&CK toolset	3
MISP	3
Adding an AMITT description to MISP	4
Disinformation object types in MISP	5
Adding an object (tweet etc) to MISP by hand	5
Adding an object to MISP via Slack bot	6
Cortex Analysers	7
Slack to MISP bots	7
Adding New Object Types to MISP	7
Disinformation object categories in MISP	8
Disinformation relationship types in MISP	9
Large Data Storage Tools	10
DKAN	10
Analysis Tools	10
Gephi	10
Viewing networks with Gephi	10
Python scripts	11
Other analysis tools	11

Response Toolkit

The tools you need depend on the size of the response you're planning to run, the number of people involved, and things like whether they already have access to their own specialised tools for things like tracking disinformation narratives. A good basic set of tools will include:

- Incident tracking. The team needs to know which incidents it's responding, where to find information on them, where to add information, and what types of action have been taken already. Incident tracking tools range from a shared spreadsheet (e.g. googlesheets and airtables) to ticketing systems (The League uses D3PO), and case management systems like TheHive.

- Incident note and summary sharing. Shared notebooks (e.g. Googledoc templates) work for this, and some tracking systems (e.g. TheHive) also include shared notes.
- Artefact analysis. We're often starting investigations from single artefacts: text, images, video, domains, groups. We borrow heavily from OSINT toolkits to analyse each of these.
- Social media analysis. At a minimum, you'll need network and text analysis tools. Some of our teams bring their own; otherwise sourcing or creating open-source analysis tools is a good thing.
- Incident technique, artefact and narrative sharing. Techniques, artefacts and narratives are objects of specific importance to an incident: they're the objects that you want to share with responders, like hashtags, groups, and superspreaders account ids. Each incident is built on techniques, artefacts and narratives: collecting, annotating, and sharing these is an important part of the teams' work. We've tried a range of tools, from shared spreadsheets (googlesheet templates) to MISP and DKAN for this.
- Incident broadcast. One group can only do so much on its own. Most of our communications to date have been through individual connections and the cross-team tracking system inside the League, but MISP allows for both setting an event to public share, and for emailing event summaries out to a subscriber list. Other possibilities include a public list of non-sensitive incidents, an incidents mailing list etc.
- Large dataset storage. A tracking team will collect a lot of supporting data that isn't artefacts: things like the tweets and accounts associated with a hashtag, or urls and groups that a story appears on. Most of this data isn't part of reports - it's supporting data - but still needs to be stored somewhere, for analysis. We've tried DKAN storage for json, CSV and image files, with sql for other objects of interest, and are investigating other storage methods.

The big idea here is that we can use existing open source threat intelligence tools for disinformation defence.

Incident Tracking Tools

HIVE

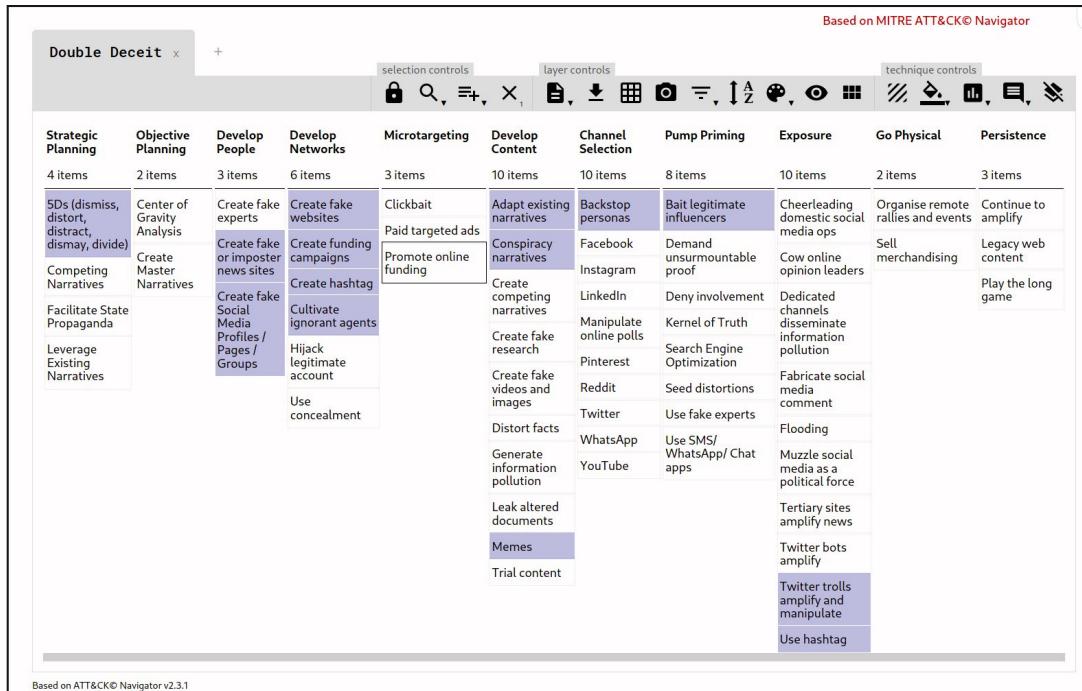
We use Hive to manage our list of incidents, and links from them to the other objects and data connected to incident responses. Check Hive and search for the incident name. All incidents will have the tag “disinformation” and word “Incident” in the title, which should help with searching.

Incident Sharing Tools

The big idea here is that we can use existing open source threat intelligence tools for disinformation defence.

For incident sharing, we've worked with the MITRE ATT&CK toolset, MISP, and OpenCTI.

MITRE ATT&CK toolset



AMITT Framework in the ATT&CK Navigator: Double Deceit Example

The AMITT Framework's ATT&CK-based format means we can reuse ATT&CK tools with it. Having STIX data was important for integration in the community, but not everyone wants to work with STIX JSON directly.

The MITRE ATT&CK navigator (image above) is well-known to most people who've used ATT&CK. It was created for navigation of STIX formatted data, is used for visualisation, red and blue team planning and has exportable layers (that can model adversary capabilities at some point in time). Tools like this are important for usability; MITRE did an excellent job on it and we hope that it will be useful to folks in the cognitive security space.

We've made a small modification to the navigator to support AMITT; this is available on the CogSec Collab site at https://www.cogsec-collab.org/project/amitt_navigator/

MISP (Malware Information Sharing Platform, https://www.misp-project.org/) is an open-source threat intelligence platform that was originally designed for malware, but is now used with many types of threat and data.

MISP is a community driven, collaborative threat intelligence platform. It's a permanent fixture of the CTI community largely due to its openness, commitment to FOSS, and an awesome community. MISP supports a range of diverse and open communities, and is used by ISAOs and ISACs, and also ad-hoc groups beyond infosec: e.g. MISP is being used to track COVID19 infections.

MISP is used to store and share structured data. It's open and extensible, and its users can easily build enrichment and automation modules for it. It also integrates with lots of other platforms and data formats including STIX. This is why we chose it for influence operations.

A good place to start with MISP is the "MISP Training for COVID" recording, from CIRCL, <https://bbb.secin.lu/b/ale-q6v-ecn>

Adding an AMITT description to MISP

AMITT Framework Galaxy interface in MISP

We built an AMITT Framework Galaxy in MISP: this now ships with MISP. A galaxy is definitions and corresponding tags, providing contextualising information. As analysts are working through reports, they can attach a technique, navigate and read a definition. For workflow, similar to the AMITT Framework, MISP allows you to click and add corresponding techniques as you work through a report.

Disinformation object types in MISP

We added a set of new object types to MISP, to help with disinformation incident tracking. Objects you might be interested in include:

Object	Misp
Facebook group	misc:facebook-group
Facebook page	misc:facebook-page
Facebook account	misc:facebook-account
Facebook post	misc:facebook-post
Twitter account	misc:twitter-account
Twitter list	misc:twitter-list
Twitter post	misc:twitter-post (was misc:microblog)
Blogsite	network:url
Blog account	misc:user-account
Blogpost	misc:blog
Reddit group (subreddit)	misc:reddit-subreddit
Reddit account	misc:reddit-account
Reddit post	misc:reddit-post
Reddit post comment	misc:reddit-comment
YouTube Channel	misc:youtube-channel
YouTube Video	misc:youtube-video
YouTube Playlist	misc:youtube-playlist
YouTube Comment	misc:youtube-comment
Website address	network:url
Hashtag	ADD NEW
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 MISP objects we might need include: 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.

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.

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.

The screenshot shows the MISP event interface for event (41): foo. At the top, there's a header "Extended by". Below it, a box contains the text "Event (41): foo" and "Currently in atomic view." with a blue circular icon. Underneath, there's a "Sightings" section with "0 (0) - restricted to own organisation only." and a small icon. At the bottom, a navigation bar includes links for Pivots, Galaxy, Event graph, Event timeline, Correlation graph, ATT&CK matrix, Attributes, and Discussion.

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>

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

Cortex Analysers

Cortex analysers are python-based tools that we can run from MISP, HIVE and Slack. We've primarily used them to speed up getting data into MISP.

Slack to MISP 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" - add a Reddit account's details
- "/misp_reddit_comment" - add a Reddit comment
- "/misp_reddit_post" - add a Reddit post
- "/misp_reddit_subreddit" - add a subreddit's details
- "/misp_builtwith" - add builtwith tags
- "/misp_twitter" - add a tweet to MISP

If we want new ones - we can build them, and Roger wrote a handy how-to guide:

<https://vx7.io/posts/2020/05/misp-slack-bot/>

Adding New Object Types to MISP

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/>

d-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)

Disinformation object categories in MISP

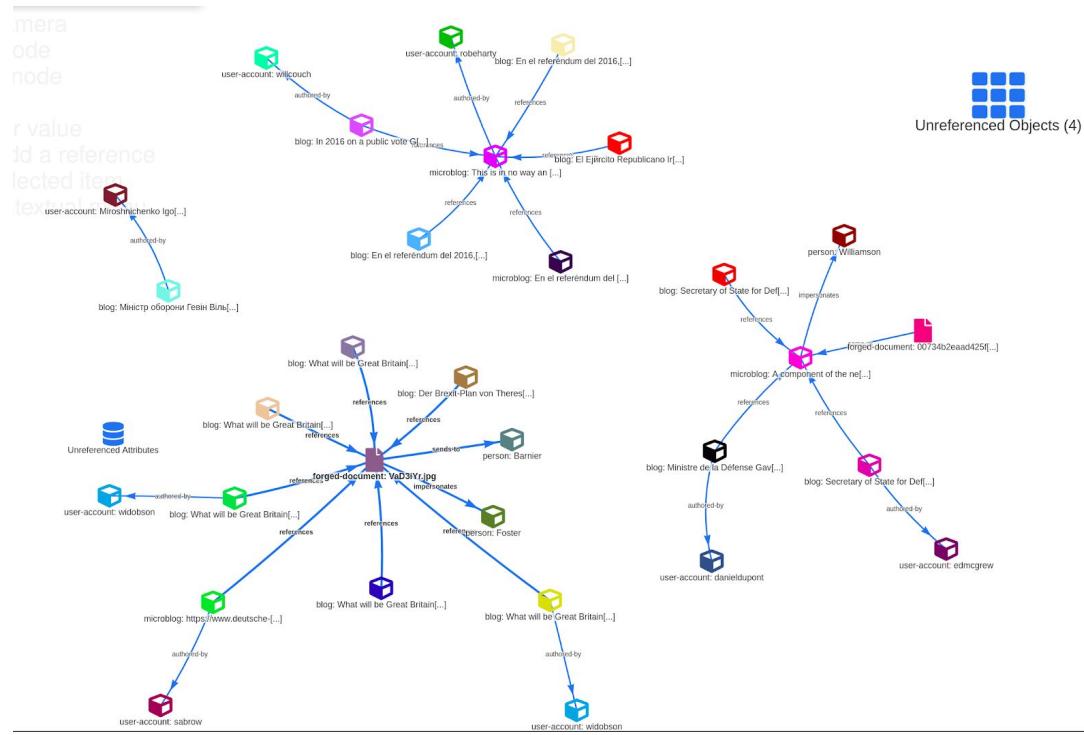
Owner org	ID	Clusters	Tags	#Attr.	Email	Date	Info
CogSec Collab	19	Misinformation Pattern	pink slime Create fake or imposter news sites Q≡	1340	info@vvx7.io	2020-01-30	U.S. Pink Slime
CogSec Collab	20	Misinformation Pattern	DFRLab-dichotomies-of-disinformation:primary-target="GB" DFRLab-dichotomies-of-disinformation:primary-disinformant="RU" DFRLab-dichotomies-of-disinformation:platforms-social-media="facebook" DFRLab-dichotomies-of-disinformation:content-topics="government" DFRLab-dichotomies-of-disinformation:content-topics="elections" DFRLab-dichotomies-of-disinformation:methods-tactics="sockpuppets" Memes Leak altered documents Tertiary sites amplify news Use hashtag Create fake Social Media Profiles / Pages / Groups Create hashtag	296	info@vvx7.io	2020-01-31	Secondary Infektion: Fantasy Assassins

MISP incident description with both DFRLab and AMITT tags

We had STIX objects in MISP for e.g. threat actors, but we don't have taxonomies for things like the types of threat actor.

We fixed this by adding the DFRLab's Dichotomies of Disinformation Taxonomy tags. This wasn't quite what we needed for tactical work, so we started working with NATO on a cutdown set.

Disinformation relationship types in MISP



MISP event graph for Sekondary Infektion

Finally, we added object-to-object relationship types to MISP to help with describing disinformation. MISP is graph-based, and these become useful when investigating and sharing relationships between objects.

The important thing is that the data we share tells a story. The AMITT framework summarises behaviour; the fancy tags (DFRlab dichotomies etc) help describe the event and provide context on what we're seeing, and MISP objects help us represent the relationships between things: Who posted a blog post, who was mentioned in a news articles, who is the registered owner of a domain etc.

Ultimately these are the things we're aiming to build and share. Not a flat list of indicators, but a model of how the adversary operated.

Large Data Storage Tools

DKAN

The screenshot shows the CogSec Collaborative DKAN interface. At the top, there's a logo for 'COG SEC COLLABORATIVE' featuring a hexagonal network icon. A search bar with a magnifying glass icon is on the right. Below the header, a navigation bar includes links for 'Datasets', 'About', 'Groups', 'Stories', 'Dashboards', 'Topics ▾', a user profile icon, 'sjterp', and 'Log out'. The main content area shows a dataset titled 'Covid19 cases'. It has a circular icon with a person silhouette. To the left, a sidebar for 'Cogsec Collab Data Science Team' shows a 'License' section for 'Creative Commons CCZero' with an 'OPEN DATA' button. The dataset description says 'COVID19-related disinformation cases and descriptions scraped from around the interwebs.' It lists three data sources: 'EUvsDisinfo disinformation cases, in csv form' (Scraped from https://...), 'COSMOS covid19 disinformation cases' (Disinformation cases from ...), and 'CMU IDEAS list of COVID19 narratives' (List of Covid19 narratives, scraped from...). Each source has 'Preview' and 'Download' buttons.

CogSecCollab DKAN

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

Analysis Tools

Gephi

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=AqIT0khVuZA

- 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.

Python scripts

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

- Learn python the hard way
- ACTION: SJ add notes on python and data science - Pablo-level friendly

Other analysis 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 17 References

Bedtime Readings	1
Books	1
Articles	1
Podcasts and videos	2
People to Follow	2
Examples of disinformation tracking	3

Bedtime Readings

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:

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

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)

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 \(Grugg, SJ, Brian\)](#)

People to Follow

Disinformation data science:

- Conspirador Norteno and Dr ZQ: [@conspirator0](<https://twitter.com/conspirator0>)[@ZellaQuixote](<https://twitter.com/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>](<https://t.co/67lk3ewfRx?amp=1>) and python/jupyter with libraries pandas, tweepy, bokeh, cytoscape
 - Looking at a botnet:
<https://twitter.com/conspirator0/status/1265829648056954881?s=20>
- Andy Patel: [@r0zetta](<https://twitter.com/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://github.com/r0zetta/meta_embedding_clustering))
 - (<https://blog.f-secure.com/author/andrew-patelf-secure-com/page/2/>)

- Elliot Alderson: [@fs0c131y](<https://twitter.com/fs0c131y>)
([fs0c131y.com](<https://t.co/ntCiqAFatC?amp=1>))
 - Infosec and misinformation data scientist

Disinformation tracking:

- Erin Gallagher: [@3r1nG](<https://twitter.com/3r1nG>)
- @josh_emerson
- [Kate Starbird: @katestarbird](<https://twitter.com/katestarbird>)

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-plan-demic-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](#)

Annex A: Covid-related disinformation

2020: Covid19 Disinformation goes big	1
Covid19-related disinformation data feeds	1
Sites to send covid19 disinformation URLs to	1
Narratives	2
Data	2
Domains	2
Covid19-related counter-disinformation feeds	2
Reports	3
Covid19 general data feeds	3
Covid19 Narratives	3
Covid19 information feeds	4
Tracking Covid19 Disinformation outside the USA	4
Places to look for non-USA disinformation	5
Non-Covid19 disinformation and where to send it	6
Covid19 disinformation references	6
Covid19 disinformation around the world	8

2020: Covid19 Disinformation goes big

The CTI League started in response to Covid19-related infosec attacks on medical and related facilities. This work grew up as Covid19 disinformation did; but just as a lot of Covid disinformation \and misinformation\ was created and amplified around the world, so too were teams and initiatives to counter that disinformation, and to provide clear sources of information on rapidly changing physical, social and scientific literature environments.

Covid19-related disinformation data feeds

Sites to send covid19 disinformation URLs to

- NewsGuard's [Coronavirus Misinformation Tracking Center](#)

Narratives

- [Wikipedia list of Covid19 rumours](#)
- [WHO Covid19 myths list](#) - narratives
- CMU IDEAS Center [list of Covid19 disinformation narratives](#) (click dates)

Data

- Ryerson University covid19 misinformation portal: <https://covid19misinfo.org/>
 - Botwatch dashboard <https://covid19misinfo.org/botwatch/>
- Uni Arkansas COSMOS Covid19 list <http://cosmos.ualr.edu/misinformation>
- Facebook datafeed: [Enabling study of the public conversation in a time of crisis](#)

Covid19-related counter-disinformation feeds

- Ryerson University covid19 misinformation portal: <https://covid19misinfo.org/> and <https://covid19misinfo.org/fact-checking/covid-19-fact-checkers/>
- 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>

Reports

- ASPI: <https://www.aspi.org.au/report/covid-19-disinformation>

Covid19 general data feeds

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

Covid19 Narratives

The CTI League concentrated on Covid19-related misinformation, and these are the top-level themes for the 100s of narratives that we saw.

- Covid isn't serious
 - Covid doesn't exist
 - Individual medical targets
- Medical scams
 - MMS prevents
 - Alcohol prevents
 - etc
- Origin myths
 - Escaped bioweapon
 - Country x created Covid
 - US soldiers took Covid to China
- Resolution myths
 - Country y has a Covid cure
- Crossover with conspiracies
 - Covid and 5G
 - Covid and antivax / anti-Gates
 - Helicopters spraying for covid
 - Depopulation conspiracy
- Crossover with 'freedom rights'
 - Anti- stayathome
 - 2nd amendment
 - anti-immigration
 - Usual far-rightwing groups
- Geopolitics
 - China, Iran: covert + overt
 - "Blue check" disinfo

Narratives are useful because there are a lot fewer narratives than messages, making them easier (especially with text-based machine learning techniques) to track. In 2020, we saw a lot of crossover narratives, where existing groups like antivaxxers and hardcore rightwingers met and joined forces, and conspiracy theories including black helicopters and 5G were recycled and combined into new Covid narratives. Most of this was worth money.

Covid19 information feeds

- <https://explaincovid.org/>

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](<https://www.cbc.ca/radio/ideas/how-hollywood-became-the-unofficial-propaganda-arm-of-the-u-s-military-1.5560575>). 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.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

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](#)

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)

Annex B: Election-related disinformation

References

1

References

General references

- Playbooks:
 - <https://www.belfercenter.org/D3P/#!the-election-influence-operations-playbook>
 - <https://cyber.harvard.edu/publication/2020/us-elections-disinformation-tabletop-exercise-package>
- Background reading
 - <https://www.ifes.org/> - election monitoring worldwide

2020 US Presidential elections

- Readings:
 - <https://about.fb.com/news/2020/08/research-impact-of-facebook-and-instagram-on-us-election/>
 - <https://www.eipartnership.net/news/twitters-policy-election-misinfo-in-action>
 - <https://go.recordedfuture.com/hubfs/reports/cta-2020-0903.pdf>
 - <https://twitter.com/2020Partnership/status/1294045600619454466?s=20>
 - <https://www.atlanticcouncil.org/event/are-we-ready-foreign-interference-disinformation-and-the-2020-election/>
 - <https://register.gotowebinar.com/register/6278844022176982285>
- Groups:
 - Election Integrity Partnership
- Background data
 - <https://verifiedvoting.org/verifier/>