Finding ways to Counter Disinformation Campaigns

Methods, mitigations and links to TTPs

=== draft 2020-01-04 ===

Contributors:

- SJ Terp
-

Contents:

Executive Summary

This report looks at existing and potential disinformation countermeasures and mitigations. It's part of a series of work on how information security principles and practices can be used to improve our understanding of and responses to disinformation campaigns and incidents.

State actors, private influence operators, and grassroots groups are exploiting the openness and reach of the Internet to manipulate populations at a distance. This is an extension of a decades-long struggle for "hearts and minds" via propaganda, influence operations, and information warfare. Computational propaganda fueled by AI has the prospect of making matters much worse.

The Credibility Coalition's MisinfoSec Working Group (MisinfosecWG) is creating standards for sharing information about misinformation incidents and how to respond to them. The work of the group is inspired largely by existing standards in information security.

The structure and propagation patterns of misinformation attacks have many similarities to those seen in information security and computer hacking. By analyzing similarities with information security frameworks, MisinfosecWG gives defenders better ways to describe, identify and counter misinformation-based attacks. Specifically, we place misinformation components into a framework commonly used to describe information security incidents. Our work will give responders the ability to transfer other information security principles to the misinformation sphere and to plan defenses and countermoves.

The report was written for the Credibility Coalition's Misinfosec Working Group, with inputs from the misinfosec community including experts who generously gave up 2 days of their time to workshop potential counters.

1 Introduction

State actors, private influence operators, and grassroots groups are exploiting the openness and reach of the Internet to manipulate populations at a distance. This is an extension of a decades-long struggle for “hearts and minds” via propaganda, influence operations, and information warfare. Recent advances include computational propaganda: the use of algorithms, including machine learning and artificial intelligence, in online manipulation.

The Credibility Coalition’s MisinfoSec Working Group (MisinfosecWG) brainstormed, collated and devised new ways to counter or mitigate online manipulation, focusing on manipulation through disinformation and its known and potential countermeasures and mitigations. This report describes that work, and is part of a series of reports on how information security principles and practices can be used to improve our understanding of cognitive security and improve responses to information operations, and specifically to disinformation campaigns and incidents.

1.1 Misinfosec

Back in 2018, the Credibility Coalition’s MisinfoSec Working Group (MisinfosecWG) started creating information-security-inspired standards for sharing information about disinformation incidents. In late 2019, they extended that standards work to how to respond to disinformation incidents and other influence operations. In 2020, MisinfosecWG joined with its sister group Misinfosec, to become CogSec Collab.

The work of these groups is rooted largely in existing work in information security. The structure and propagation patterns of misinformation attacks have many similarities to those seen in information security and computer hacking. By analyzing similarities with information security frameworks, MisinfosecWG gave defenders better ways to describe, identify and counter misinformation-based attacks. Specifically, we placed misinformation components into a framework, AMITT, based on standards (including ATT&CK and STIX) commonly used to describe information security incidents.

Our intent was always to give responders the ability to transfer other information security principles to the misinformation sphere, and to plan defenses and countermoves. We started the disinformation countermeasures work with a two-day workshop in Washington DC, with two main goals:

- Create the first version of a disinformation “Blue Team” playbook. For defenders, information security people and organizations, this will be a set of responses to misinformation attacks—the networks, the response types, the frameworks, and examples.

- Define how to support an operational global MisinfoSec_ISAO network. For potential response center participants and leaders, this will be the process, methods and understanding needed to connect, including suggesting partners, collaborators and funders.

This report is one of the first outputs from the workshop.

1.2 Definitions

There are many definitions of misinformation, disinformation, incident etc. and teams dedicated to improving them. We don't want to get into that field, so for this report, we've used a set of working definitions. These are:

-

<action: define misinformation, disinformation, countermeasure, and any other terms in here that might be confused>

2 Finding Countermeasures

2.1 Introduction

The AMITT framework so far is a beautiful thing — we've used it to decompose different misinformation incidents into stages and techniques, so we can start looking for weak points in the ways that incidents are run, and in the ways that their component parts are created, used and put together. But right now, it's still part of the "admiring the problem" collection of misinformation tools -to be truly useful, AMITT needs to contain not just the breakdown of what the blue team thinks the red team is doing, but also what the blue team might be able to do about it. Colloquially speaking, we're talking about countermeasures here.

There are several ways to go about finding countermeasures to any action:

- Look at counters that already exist. We've logged a few already in the AMITT repo, against specific techniques — for example, we listed a set of counters from the Macron election team as part of incident I00022.
- Pick a specific tactic, technique or procedure and brainstorm how to counter it — the MisinfosecWG did this as part of their Atlanta retreat, describing potential new counters for two of the techniques on the AMITT framework.
- Wargame red v blue in a 'safe' environment, and capture the counters that people start using. The Rootzbook exercise that Win and Aaron ran at Defcon AI Village was a good start on this, and holds promise as a training and learning environment.
- Run a machine learning algorithm to generate random countermeasures until one starts looking more sensible/effective than the others. Well, perhaps not, but there's likely to be some measure of automation in counters eventually…

2.2 Searching for Countermeasures

Searching for disinformation resources at the end of 2019 is much easier than in previous years. Major lists of projects, reports and groups that yielded existing countermeasures included

- Oxford Internet Institute's computational propaganda project's resource finder https://navigator.oii.ox.ac.uk/resources/?resource_filter%5Bsubject%5D%5B%5D=disinformation-counter-strategies#
- Rand.org's reports on disinformation (e.g. [Rand2740])
- Scott Yate's CCC lists of projects, and the Credibility Coalition's navigator

Many other groups (CMU etc) are creating their own lists, making this a great time to hunt for specific counters.

2.3 Known Countermeasures

There are many published “solutions” to disinformation attacks. While useful, it’s foolish to consider any of these the “silver bullet” that solves a disinformation problem; they often address smaller pieces of an attack, or are intractable or don’t scale. Disinformation campaigns are whole-system attacks: to solve them we need to look at whole-system solutions: this is more of a “thousand bullet” solution than a single-bullet one. Some components in the current counter landscape are:

- Detecting artificial amplification. Many disinformation campaigns rely on signal amplification, either through ‘useful idiots’ or by raising message visibility using non-human traffic (‘bots’ and ‘botnets’). Databases of known online bad actors and state-sponsored actors, with data from pages and social media feeds from these actors have proven useful places to look for emerging narratives and links to new actors and artefacts. Tracking bots and botnets has become more difficult as adversaries adapt to detection techniques (both from disinformation detection but also from adjacent domains including mitigating advertising click fraud) and trade message reach for keeping valuable networks online, but there is still value in simple bot/botnet detection techniques including analysis of similarities across accounts linked by topic, hashtags, retweets, references etc, and time-series analysis to check for sleep/wake patterns, activity correlations etc, especially with adversaries new to this space.

- Detecting related artifacts. Disinformation campaigns rarely use one account, platform, account network or domain, and financially-motivated campaigns sometimes run sets of sites with wildly different topics or demographic/country targets. Most work on this isn’t tool-based; it’s digital forensics, tracking artifacts like tag IDs, domain registrations and reused/linked content across the internet using OSINT tools (Bellingcat and DigitalSherlocks both publish good examples of this work).

- Mitigating artificial amplification. Most current work on this is platform takedowns or “shadow-banning” of known bot, botnet, troll or other artificial amplification social media accounts. Related work includes removal of online advertising and product revenue from domains that are part of financially-motivated disinformation campaigns.

- Resilience against adversarial narratives. It’s preferable to remove a disinformation campaign before it reaches the general population, but if it does, building resilience to disinformation campaigns in the form of awareness of techniques, critical reasoning skills etc is useful. Most population resilience counters are in the form of education - either at school level or through information campaigns like the US State Department’s War on Pineapple posters. More active population resilience measures include the Baltic Elves volunteer groups posting disclaimers and counter-narratives to Russian disinformation in their countries.

2.3.1 Nationstate counters

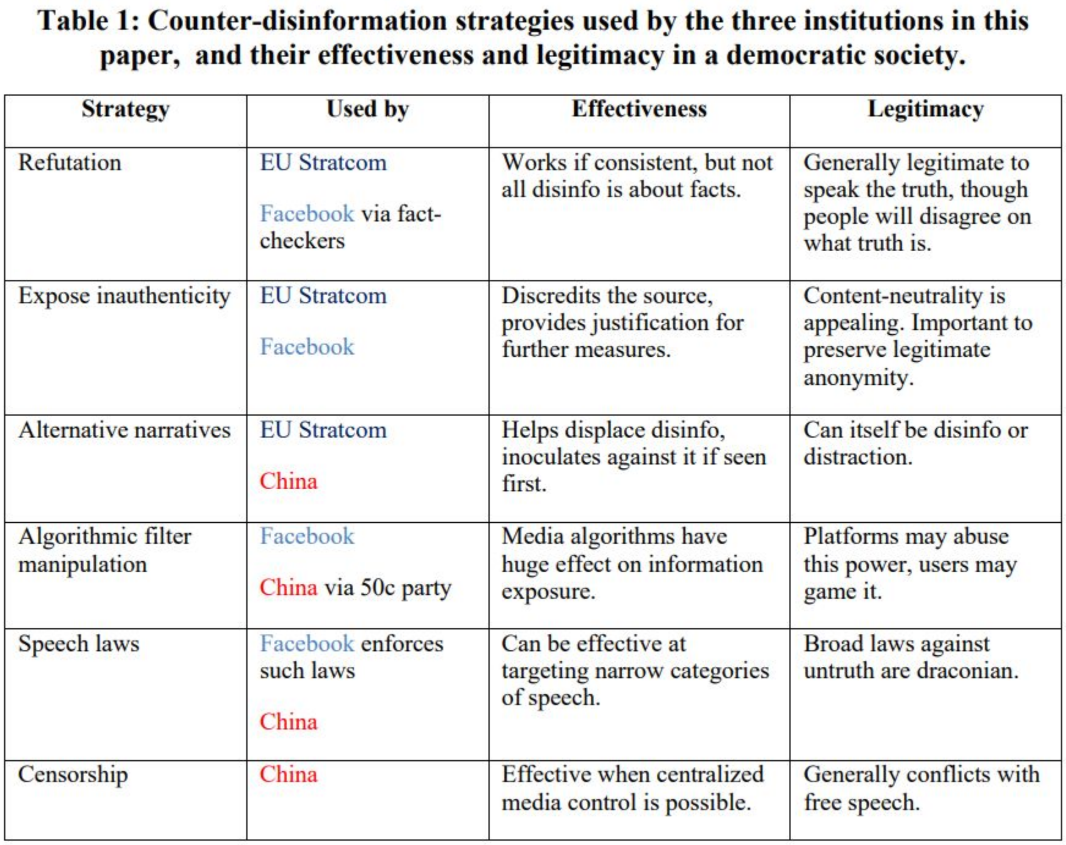


Table 1: Counter-disinformation strategies used by the three institutions in this paper, and their effectiveness and legitimacy in a democratic society.

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.

“A taxonomy of tactics” from [Stray19]

2.4 Countering AMITT components

Work on AMITT used existing information security models (e.g. cyber kill chain, ATT&CK) to model disinformation incidents as collections of tactics, techniques and procedures (TTPs). One way to look at counters is to look at that breakdown and find or devise responses and mitigations to each TTP. At the tactic level, this gives us a Courses of Action matrix (COA), with the tactic stages listed on one axis, and types of response - eg. (Deny, Disrupt, Degrade, Destroy) - on the other, At the technique level, this gives us a way to discuss mitigations for techniques (e.g. the use of botnets) that we see repeatedly in disinformation incidents.

This is one way to look at countermeasures and mitigations. It’s a useful way to examine the space of possible actions, in the same way that a naval officer learns about ‘standard’ manoeuvres like the Crazy Ivan, and how to think about detecting and mitigating for them. Disinformation, like war, isn't a linear process: that there are techniques in play that work and are likely to be used is just the first level of understanding what could and might be done. Good incident creators are also artists (yes, yes, there's a reason it's called "the Art of War"), understanding the basic techniques and constraints, and knowing how to adapt them into a flow of actions that becomes difficult to counter with a simple rulebook. These masters still need to know the basics though.

2.5 2.4Workshopping Counters

Day 1
● Introduce what MisinfoSec_WG has done, why we've done it, and what we have to show. Introduce AMITT; review stages and techniques
● Workshop/hands on "Blue Team" to build the responses part of the framework
● Create 5-7 five-person multi-disciplinary teams each responsible for creating a collection of counters for up to 10 of the 54 identified techniques

Day 2
● Introduce ISAO concepts and how they connect to misinformation
● Workshop/hands on design of ISAO network support
● Workshop/hands on exercise testing responses and network concept together
● Wash-up and next steps

2.6 Building Playbooks

A big pile of countermeasures is nice, but it's not going to help someone who's facing an immediate active campaign or incident. They're going to need some form of "hey, this is happening, here are things you could try and what might happen" guides.

When organising countermeasures, there are a few questions to ask:

- What does this counter? Is this a mitigation, and what does it do: does it stop a technique being effective, moderate its effect or do something like delay its effect whilst other measures are put in place?
- Who can do this? What skills and resources do they need to have a chance at success? What risks do they take in doing it and how can those be both explained and minimised?
- Has this been tried before? What happened that time? Are there side effects (both good and bad) to watch out for?
- Has this been used in combination with other counters? Could it be?
- What level is this counter at? Is it strategic, tactical or immediate?

3 Making Countermeasures Easy to Share

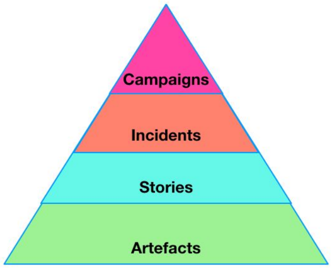
Online disinformation doesn't exist in a vacuum. The same types of framework that help campaign creators can also help with their removal. For instance, the easy access to demographic datasets that make micro targeting easy could be countered with stronger use of privacy laws and individual counters against online privacy invasions.

3.1 Disinformation as an Ecosystem

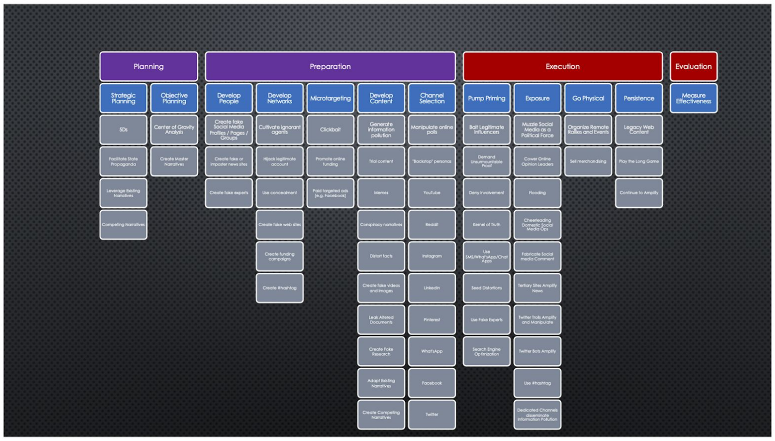
One of the things that reading through the counters spreadsheet surfaces is the sense of who is doing what to whom with which resources? For example - we have a lot of entries that look something like "tell x about y". Which is great, but that assumes that y can do something about x. After a while this starts to look like pieces of a stix graph itself - we have actors (or types of actor), artefacts and techniques in play, connecting to and relying on each other. Content takedowns, for instance: these can only happen if the people capable of doing the takedowns know about the content, and the people who know about the content tell them about it. We may also have a componentwise, piece-together set of responses to be built. To start with, mapping out who is doing what to whom with which resources, and which assumptions about actions and outcomes might go a long way in reducing our 200+ listed counters down to a manageable tactical set.

3.1.1 Component-based views of disinformation

We can look at the solution space in several different ways. One of them is as a human space, in which we're engaged in narrative warfare. Human communication is generally at the level of stories, or narration: we tell each other stories about the world, as sentences, image sequences, or memes. Narratives are the stories that each person and community bases their sense of self, their belonging to different groups ("in-groups"), and exclusion of others ("out-groups") on. Narratives are typically personal, emotionally-charged, deeply-entrenched and difficult to shift directly. In this space, it becomes important to track narratives and their components (e.g. memes, stories, sentiments) and disrupt them not by countering them directly with 'facts', but with 'information aikido': it's easier to redirect an angry mob to a different house than it is to disband them. Narrative warfare is a growing field, and its techniques are a useful component in countering disinformation. Using Natural Language Processing techniques like topic modelling and gisting to track narratives from disinformation actors, and highlighting narratives to potential target audiences have also proved useful.

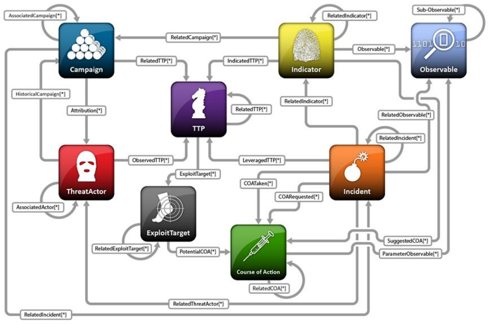


The misinformation pyramid is another view of this space. Here we're looking at the different views that creators of misinformation ('attackers') and the people trying to counter them (‘defenders’) have (the third group involved, the targets of the misinformation, ‘populations’, aren’t part of this diagram). Attackers create incidents (e.g. Macrongate), which often form part of longer-term campaigns (e.g. destabilize French politics). Narratives are the stories that we base our beliefs on. To tell these stories, we need artifacts: the users, tweets, images, etc., that are visible in each attack. While 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.

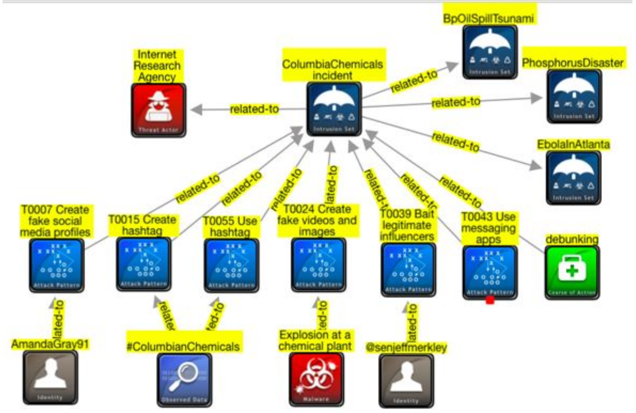


Planning		Preparation					Execution				Evaluation
Strategic Planning	Objective Planning	Develop People	Develop Networks	Microtargeting	Develop Content	Channel Selection	Pump Priming	Exposure	Go Physical	Persistence	Measure Effectiveness
5Ds	Center of Gravity Analysis	Create fake Social Media Profiles / Pages / Groups	Cultivate ignorant agents	Clickbait	Generate information pollution	Manipulate online polls	Bait Legitimate Influencers	Muzzle Social Media as a Political Force	Organize Remote Rallies and Events	Legacy Web Content	
Facilitate State Propaganda	Create Master Narratives	Create fake or imposter news sites	Hijack legitimate account	Promote online funding	Trial content	"Backstop" personas	Demand Unsurmountable Proof	Cower Online Opinion Leaders	Sell merchandising	Play the Long Game	
Leverage Existing Narratives		Create fake experts	Use concealment	Paid targeted ads [e.g. Facebook]	Memes	YouTube	Deny Involvement	Flooding		Continue to Amplify	
Competing Narratives			Create fake web sites		Conspiracy narratives	Reddit	Kernel of Truth	Cheerleading Domestic Social Media Ops			
			Create funding campaigns		Distort facts	Instagram	Use SMS/What'sApp/Chat Apps	Fabricate Social media Comment			
			Create #hashtag		Create fake videos and images	Linkedin	Seed Distortions	Tertiary Sites Amplify News			
					Leak Altered Documents	Pinterest	Use Fake Experts	Twitter Trolls Amplify and Manipulate			
					Create Fake Research	What'sApp	Search Engine Optimization	Twitter Bots Amplify			
					Adapt Existing Narratives	Facebook		Use #hashtag			
					Create Competing Narratives	Twitter		Dedicated Channels disseminate Information Pollution			

A useful view of a disinformation incident is as a collection of the objects seen within it, including the techniques, tactics and procedures (TTPs) that the adversary used. The AMITT model (based on the ATT&CK framework) describes common disinformation TTPs across 12 stages of adversary activity, from strategic planning of each incident to evaluating its effectiveness and lessons learned from the deployment, as a feed into later incident plans.



This gives us a third view of the space. Disinformation incidents are rarely isolated events. TTPs are part of richer description languages including STIX, that allow analysts to share and compare information about shared threat actors, narratives, TTPs, artifacts and other objects in each incident and campaign.



With this view, we can start to record and recall previous countermeasures to reused techniques, and find and exploit weaknesses and gaps in the adversary's operations, in the same way we exploit adversary weaknesses in gaps in other situation pictures, including those in cybersecurity.

3.2 Component based views of response

Adversary tactics are moving quickly in this arena (for instance, we're just starting to see the types of tool changes seen in the related field of MLsec), so tools and counter tactics are likely to change but the basic problems won't. Rather than enumerating countermeasures and tools, it's useful to examine the disinformation solution space, and consider the tools and techniques that might be needed in it.

3.2.1 Courses Of Action Matrix

If we're following the infosec playbook to do all this faster this time around (we really don't have 20 years to get decent defences in place — we barely have 2…) then shouldn't we look at things like courses of action matrices? Yes. Yes we should…

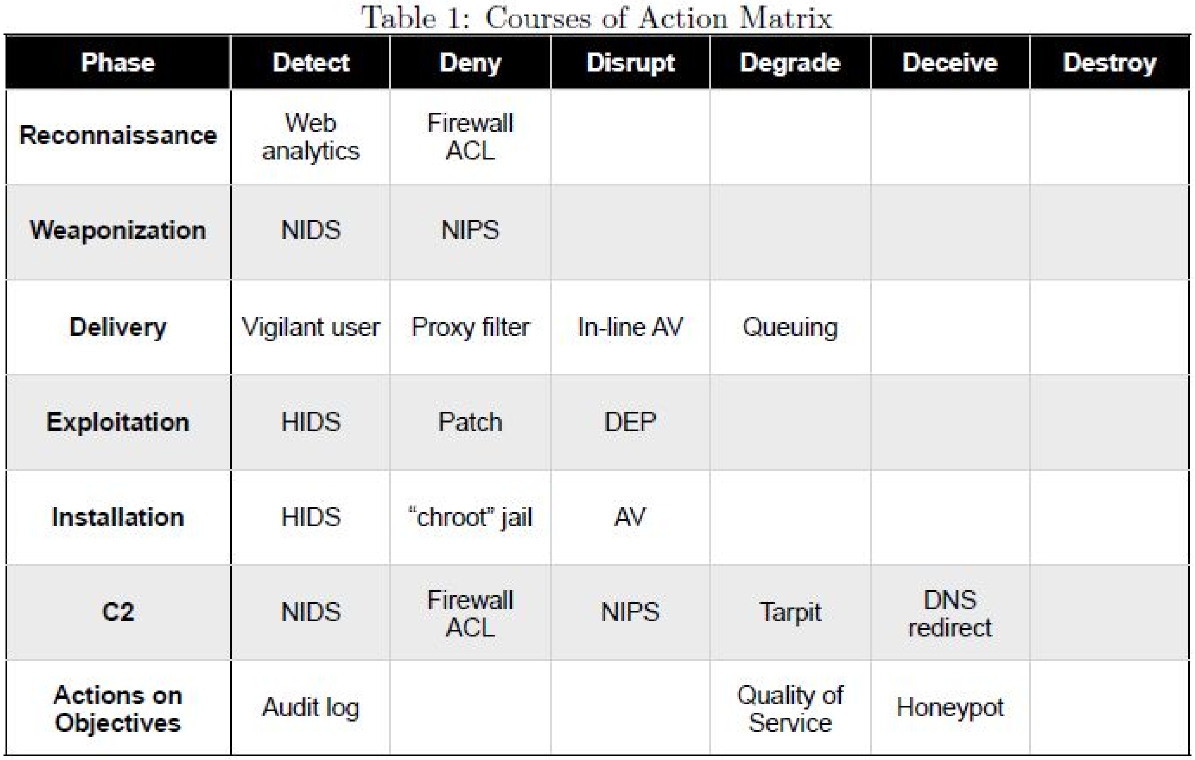


Table 1: Courses of Action Matrix

Phase	Detect	Deny	Disrupt	Degrade	Deceive	Destroy
Reconnaissance	Web analytics	Firewall ACL				
Weaponization	NIDS	NIPS				
Delivery	Vigilant user	Proxy filter	In-line AV	Queuing		
Exploitation	HIDS	Patch	DEP			
Installation	HIDS	"chroot" jail	AV			
C2	NIDS	Firewall ACL	NIPS	Tarpit	DNS redirect	
Actions on Objectives	Audit log			Quality of Service	Honeypot	

Courses of Action Matrix [1]

So this thing goes with the Cyber Killchain — the thing that we matched AMITT to. Down the left side we have the stages; 7 in this case; 12 in AMITT. Along the top we have six things we can do to disrupt each stage. And in each grid square, we have a suggestion of an action (I suspect there are more than one of these for each square) that we could take to cause that type of disruption at that stage. That's cool. We can do this.

The other place we can look is at our other parent models, like the ATT&CK framework, the psyops model, marketing models etc, and see how they modelled and described counters too — for example, the mitigations for ATT&CK T1193 Spearphishing.

Action: add Anti-harassment models

3.2.2 Hybrid responses

There is no one, magic, response to misinformation. Misinformation mitigation, like disease control, is a whole-system response.

MisinfosecWG has been working on infosec responses to misinformation. Part of this work has been creating the AMITT framework, to provide a way for people from different fields to talk about misinformation incidents without confusion. We're now starting to map out misinformation responses, e.g.

- At the technique level — T0025 leak altered documents was countered in France during the Macron election.

- At the tactic level — we can create a courses of action matrix that lists ways to detect, deny, disrupt, degrade, deceive or destroy activities in each tactic stage.
- At the procedure level — we can look at sequences of responses that may be more effective than individual responses in isolation.

The Rootzbook misinformation challenge take shape. I'm impressed at what the team has done in a short period of time (and has planned for later). It also has a place on the framework — specifically at the far-right of it, in TA09 Exposure. Other education responses we've seen so far include:

- Immunisation through gameplay "pre-bunking", e.g. the game https://getbadnews.com/#intro
- Education on specific techniques, e.g. the pineapple pizza education on division tactics
- The Finnish education model
- Other counters being explored by groups like the CredCo media literacy working group

Education is an important counter, but won't be enough on its own. Other counters that are likely to be trialled with it include:

- Tracking data providence to protect against context attacks (digitally sign media and metadata in a way that media includes the original URL in which it was published and private key is that of the original author/publisher)
- Forcing products altered by AI/ML to notify their users (e.g. there was an effort to force Google's very believable AI voice assistant to announce it was an AI before it could talk to customers)
- Requiring legitimate news media to label editorials as such
- Participating in the Cognitive Security Information Sharing and Analysis Organization (ISAO)
- Forcing paid political ads on the Internet to follow the same rules as paid political advertisements on television
- Baltic community models, e.g. Baltic "Elves" teamed with local media etc

Jonathan Stray's paper "Institutional Counter-disinformation Strategies in a Networked Democracy" is a good primer on counters available on a national level.

3.3 Sharing Formats

Checking parent models is also useful because this gives us formats for our counter objects— which is basically that these are of type "mitigation", and contain a title, id, brief description and list of techniques that they address. Looking at the STIX format for course-of-action gives us a similarly simple format for each counter against tactics — a name, description and list of things it mitigates against.

We want to be more descriptive whilst we find and refine our list of counters, so we can trace our decisions and where they came from. A more thorough list of features for a counter list would probably include:

- id
- name
- brief description
- list of tactics can be used on
- list of techniques can be used on
- expected action (detect, deny etc)
- who could take this action (this isn't in the infosec lists, but we have many actors on the defence side with different types of power, so this might need to be a thing)
- anticipated effects (both positive and negative — also not in the infosec lists)
- anticipated effort (not sure how to quantify this — people? money? hours? but part of the overarching issue is that attacks are much cheaper than defences, so defence cost needs to be taken into account)

And be generated from a cross-table of counters within incidents, which looks similar to the above, but also contains the who/where/when etc:

- id
- brief description
- list of tactics it was used on
- list of techniques it was used on
- action (detect, deny etc)
- who took this action
- effects seen (positive and negative)
- resources used
- incident id (if known)
- date (if known)
- counters-to-the-counter seen

4 Coordinating Responses

We need to tie this all together. Whole-system attacks often need whole-system responses. We've seen campaign creators different types of account (bot, troll, cyborg, 'useful idiots' etc) across multiple platforms, topics and geographies.

4.1 Who can act and how?

Describing actions is great, but actions only work if someone does them. There are many entities in the space of being affected by and analysing disinformation campaigns; not so many entities in the space of being able to, willing to, legally allowed to, or actively responding to disinformation.

Entities who could respond include social media platforms, other organisations, civil society, media organisations, governments, militaries and individuals. There are also other stakeholders who could be persuaded or find it in their best interests to help reduce the prevalence of disinformation campaigns across societies.

Social media platforms have control over their own software, and usually have control over the data moving through it, and the data available on and archived in it. They also have control over who can access that software and data - or rather, over which accounts can access it. Very few social media companies are owned by individuals now - they tend to be accountable to business stakeholders whose motivation is, generally, profit. This means that removing disinformation from systems is often in competition or conflict with other business priorities, or may require system adaptations or rebuilds that are too costly to justify against an uncosted, unquantified, unknown damage to society.

Other online organizations include organizations like web hosts and DNS registrars, who could help with the removal of disinformation campaign websites.

Civil society is that connector between the people trying to help counter disinformation campaigns and the people who are subjected to them. This is where people-centre approaches like education and reporting routes for microtargeted messages and advertisements are tried.

Media has its own disinformation problems, despite its emphasis to itself on trying to find truth. Falling media budgets, longer/faster news cycles and wide access to information about breaking stories has left individual net journalists struggling to keep up and wade through streams of information, malinformation and disinformation around events. The counters here are two-way - both journalists helping counter disinformation with new practices (e.g. "rumour" pages during natural disasters), and in better training on content ingestion and dissemination practices.

Governments can help primarily with the regulations that companies can use to justify moving disinformation measures above other line items in their business plans. The shadier parts of government can also help with more direct action tracking down and dissuading campaign creators and amplifiers.

4.2 The ISAC/ISAO system

4.3 Running a Coordinating Body

Needs to feed into isaos and isacs. Needs to connect platforms and other responders. Needs a route for 'ordinary citizens' to report disinfromation artefacts.

Issues:

- Has to be neutral
  Must expect to get 'played' with adversarial data
- Has to handle commercial data to not leak commercial secrets or give undue commercial advantage
- Has to handle sensitive PII issues, whilst still alerting members to campaigns and incidents
- Has to be fast
- Has to manage a diverse, distributed community

Staffing:

- Community manager
- Data scientists / OSINT specialists
- Developers

Technologies:

- STIX-based datastore of known actors, narratives, incidents, campaigns
- TAXII server to distribute real-time messages to members
- Email newsletters to keep members up-to-date on technologies, changing trends, flash alerts etc

Datafeeds in

- Weekly "what happened this week" strategic notes - events, incidents, tools etc
- Flash alerts
- Data flows

    - https://securingdemocracy.gmfus.org/hamilton-dashboard/ - "Version 2.0 of Hamilton 68 displays outputs from sources that we can directly attribute to the Russian government or its various news and information channels" - raw artefacts
    - Omelas - https://www.crunchbase.com/organization/omelas - raw artefacts on russian, china, iran etc originated
    - Marvelous.ai tracking narratives around specific events (e.g. political debates)
    - NATO StratcomCOE tracking eastern european narratives / incidents https://www.facebook.com/pg/StratComCOE/posts/
    - Botsentinel https://botsentinel.com/ - list comes from online trollhunting

Data management

- STIX messaging and databases

Groups who can respond

- Google
- Twitter/facebook
- Smaller social media
- Salesforce/Okta/ other identity management

4.4 The limits of standards-based countermeasure approaches

At this stage, older infosec people are probably shaking their heads and muttering something about stamp collecting and bingo cards. We get that. We know that defending against a truly agile adversary isn't a game of lookup, and as fast as we design and build counters, our counterparts will build counters to the counters, new techniques, new adaptations of existing techniques etc.
But that's only part of the game. Most of the time people get lazy, or get into a rut — they reuse techniques and tools, or it's too expensive to keep moving. It makes sense to build descriptions like this that we can adapt over time. It also helps us spot when we're outside the frame.

5 Summary and Future Work

6 References

Who needs to read this section: someone going looking for more counters.

General

- [Rand2740] Bodine-Baron et al, “countering Russian social media influence”, Rand report RR 2740, 2018
- [Bradshaw18] Bradshaw et al, “Government responses to malicious use of social media”, NATO STRATCOM COE, 2018

Must-reads on counters

- [Stray19] Jonathan Stray, “Institutional Counter-disinformation Strategies in a Networked Democracy”, WWW 2019 (video)
- https://www.dhs.gov/sites/default/files/publications/19_0717_cisa_the-war-on-pineapple-understanding-foreign-interference-in-5-steps_0.pdf (war on pineapple)
- Chapter 7 of https://www.foreign.senate.gov/imo/media/doc/FinalRR.pdf

Papers to read

- AMITT main paper; summary of AMITT main paper
- [Darczewska14] Jolanta Darczewska, “The Anatomy of Russian Information Warfare. The Crimean operation, a Case Study”, Point of View 42, Warsaw, May 2014
- Jonathan Corpus Ong, Jason Vincent A Cabanes, “Architects of Networked Disinformation. Behind the scenes of troll accounts and fake news production in the Philippines”, Newton Tech4Dev Network, 2018 (summary)
- Keir Giles, “Handbook of Russian Information Warfare”, NATO Defense College, November 2016
- MITRE, “Getting started with ATT&CK”, October 2019
- Olga Robinson, Alistair Coleman, Shayan Sardarizadeh, “A report of anti-disinformation initiatives”, Oxford Internet Institute, August 2019

General references on counters

- Countering disinformation: three levels of action
- The ‘dark side’ of digital diplomacy: countering disinformation and propaganda
- Three Things Facebook Could Do to Suck Less
- The ASD Policy Blueprint for Countering Authoritarian Interference in Democracies
- 2018 Ranking of countermeasures by the EU28 to the Kremlin's subversion operations
- Countering information influence activities : A handbook for communicators
- Commanding the Trend: Social Media as Information Warfare
- On Cyber-Enabled Information/Influence Warfare and Manipulation

- The Russian "Firehose of Falsehood" Propaganda Model: Why It Might Work and Options to Counter It
- The Future of Political Warfare: Russia, the West, and the Coming Age of Global Digital Competition
- Analyzing the Ground Zero. What Western Countries can Learn From Ukrainian Experience of Combating Russian Disinformation - European Values Center for Security Policy
- Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States by William H. Dutton, Bianca Reisdorf, Elizabeth Dubois, Grant Blank
- Overview of countermeasures by the EU28 to the Kremlin's subversion operations: How do the EU28 perceive and react to the threat of hostile influence and disinformation operations by the Russian Federation and its proxies? 2018 Ranking of countermeasures by the EU28 to the Kremlin's subversion operations
- By Other Means Part I: Campaigning in the Gray Zone
- By Other Means Part II: Adapting to Compete in the Gray Zone
- Joint Concept Note JCN 2/18 Information Advantage
- Stemming the VIRUS: Understanding and responding to the threat of Russian disinformation
- Russia- Proofing Your Election: Global lessons for protecting Canadian democracy against foreign interference
- The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment
- Russian Social Media Influence: Understanding Russian Propaganda in Eastern Europe
- The Black Market for Social Media Manipulation
- https://ukraineelects.org/live-updates/page/4/
- Counters sections in The Cognitive Campaign: Strategic and Intelligence Perspectives
- https://open.nytimes.com/introducing-the-news-provenance-project-723dbaf07c44?gi=18ed205c8ce2
- https://www.csis.org/analysis/successfully-countering-russian-electoral-interference
- Wolfgang Schulz, "roles and responsibilities of information intermediaries. Fighting misinformation as a test case for a human rights-respecting governance of social media platforms", Hoover Institution
- Hutchins et al "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains", 2011
- Woolley https://www.economist.com/open-future/2020/01/17/digital-disinformation-is-destroying-society-but-we-can-fight-back
- https://navigator.oii.ox.ac.uk/resources/?resource_filter%5Bsubject%5D%5B%5D=disinformation-counter-strategies#
- http://verificationhandbook.com/additionalmaterial/
- https://www.poynter.org/fact-checking/2018/misinformed-podcast-who-is-fact-checking-actually-for/

- https://www.wired.com/story/opinion-to-bolster-cybersecurity-the-us-should-look-to-estonia/
- https://www.fpri.org/article/2020/02/defeating-disinformation-threats/
-

Annex: notes on the counters playbook

We included a column, ethics, in the countermeasures tab, to reflect the different ethical constraints on countermeasures, e.g. offensive moves might be inaccessible to government agencies. We haven't populated it yet, so have removed it for now.

Notes

NLP, social graph analysis, propagation patterns. Lots of approaches that are only pieces of the puzzle, or intractable/unscaleable. Difficulty of counteracting entrenched beliefs directly, information aikido, disrupting the coordination of meme/conspiracy attacks. Importance of information sharing for detecting campaigns early. AMITT, kill chains, counterterrorism models. Potential for AI/ML approaches to detection and automated countermeasures.

Links to check

- https://www.rand.org/research/projects/truth-decay/fighting-disinformation.html
- https://medium.com/@Exovera/tracking-media-narratives-275ce78d7303

Mention Kathleen Carley's work: "I would call the group's attention to the work of Dr. Kathleen Carrley of Carnegie-Mellon. She has learned to characterize the specific "information maneuvers" of adversaries in terms of specific operational actions and desired intent, and the effects on target networks resulting from these maneuvers. She terms this field "social cybersecurity." "
https://www.armyupress.army.mil/journals/military-review/english-edition-archives/mar-apr-2019/117-cybersecurity/#.Xg40eR8xcag.mailto