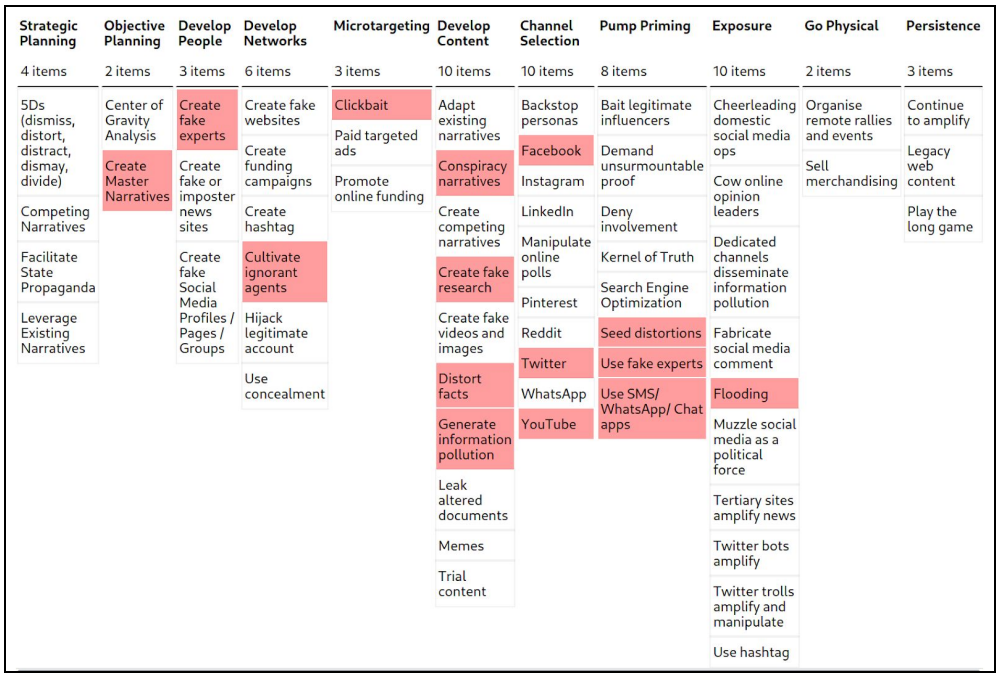
4. Disinformation Models



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
5Ds (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	Demand unsurmountable proof	Cow online opinion leaders	Sell merchandising	Legacy web content
Facilitate State Propaganda		Create fake Social Media Profiles / Pages / Groups	Create hashtag	Promote online funding	Create competing narratives	Instagram	Deny involvement	Dedicated channels disseminate information pollution		Play the long game
Leverage Existing Narratives			Cultivate ignorant agents		Create fake research	LinkedIn	Kernel of Truth	Fabricate social media comment		
			Hijack legitimate account		Create fake videos and images	Manipulate online polls	Search Engine Optimization	Flooding		
			Use concealment		Distort facts	Pinterest	Seed distortions	Muzzle social media as a political force		
					Generate information pollution	Reddit	Use fake experts	Tertiary sites amplify news		
					Leak altered documents	Twitter	Use SMS/ WhatsApp/ Chat apps	Twitter bots amplify		
					Memes	WhatsApp		Twitter trolls amplify and manipulate		
					Trial content	YouTube		Use hashtag		

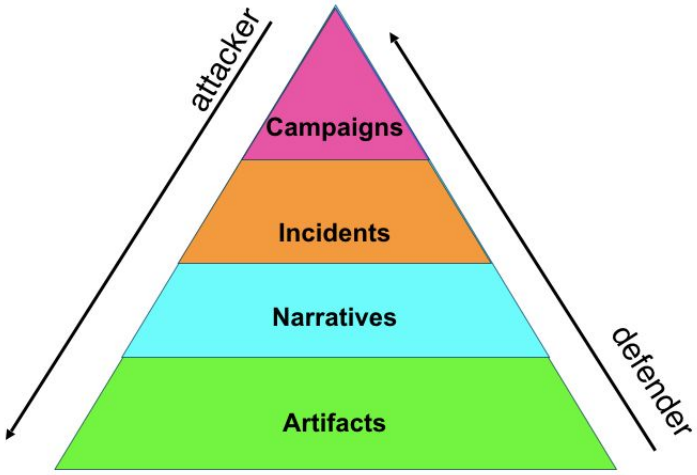
TL;DR Disinformation Models

- An incident is a coordinated set of activities, over a relatively-short timespan, usually with an individual or team behind it.
- We're using adapted information security standards to describe disinformation incidents, so we can share them with a large number of responders.
- We describe incidents 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.
- 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

Disinformation Models

Models help us understand and share information about disinformation. Models also help us plan misinformation defenses and counters, assess tools and mechanisms, and handle adaptive threats created with machine learning.

Layer Models



Disinformation Pyramid

The disinformation pyramid connects 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.

Actor Models

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

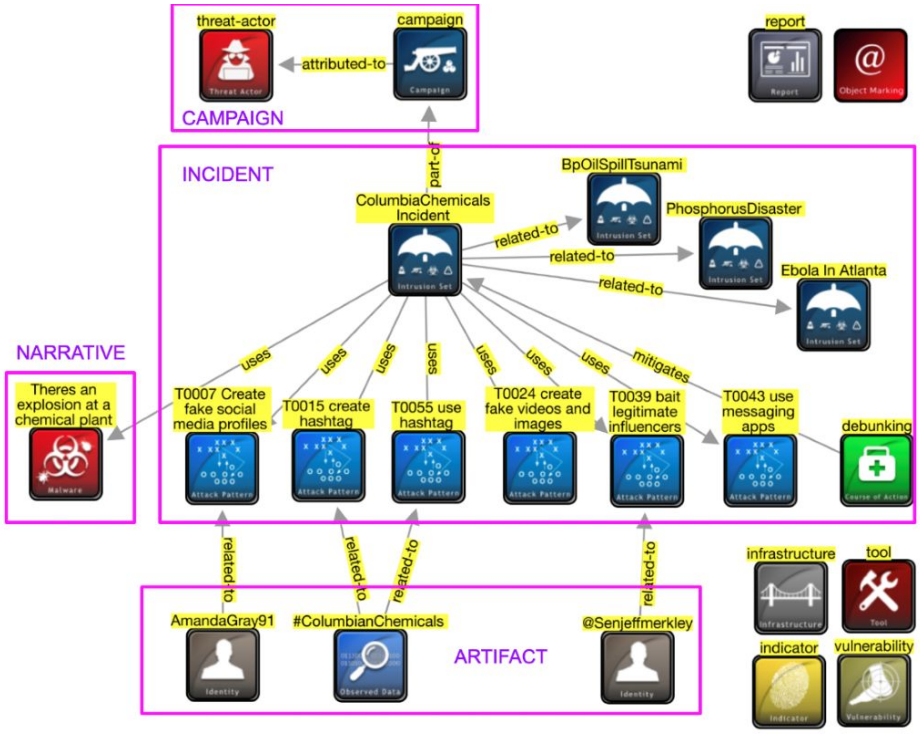
to accomplish and design a misinformation campaign to achieve that goal. They manufacture one or more incidents, each incident has its own narrative which is told through a series of artifacts. Those artifacts can be posts, tweets, stories, deep fakes, etc. As attackers move down the pyramid, more work 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 ever 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.

Object Models

AMITT STIX



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	Artefact	Observed Data
Indicator	posting rates, follow rates etc	Artefact	Indicator
Vulnerability	Cognitive biases, community structural weakness etc	Vulnerability	Vulnerability

Disinformation version of STIX

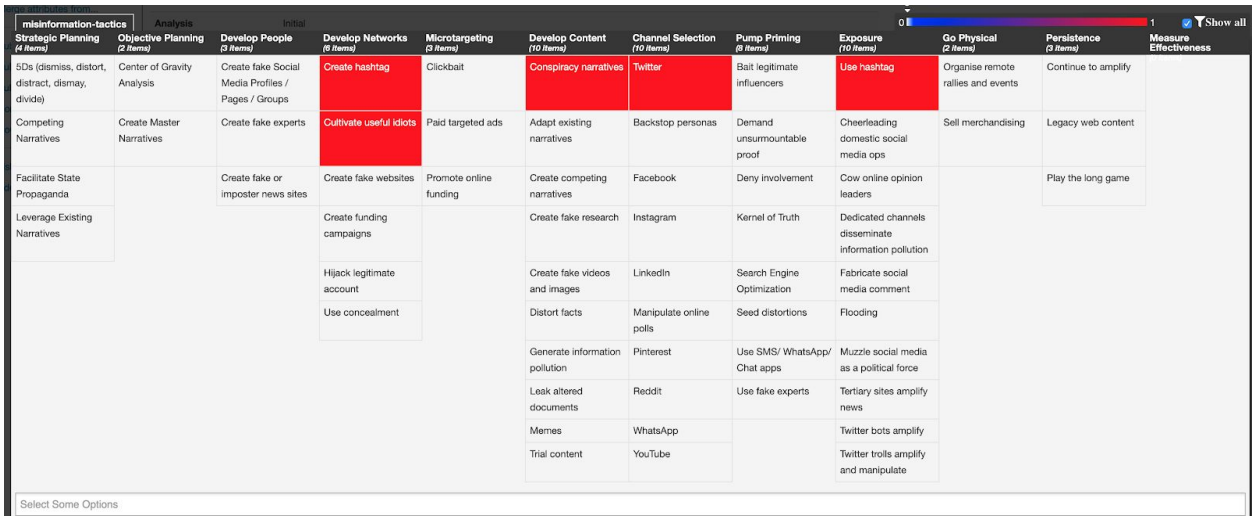
STIX translates well for disinformation use. We added two objects to STIX for disinformation: incident, and narrative, and didn't need to change anything else. We use custom objects to represent these fields and be OpenCTI compliant.

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.

Behaviour Models

Disinformation TTPs: Tactics, Techniques, Procedures



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

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 AMITT Github repository - 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.

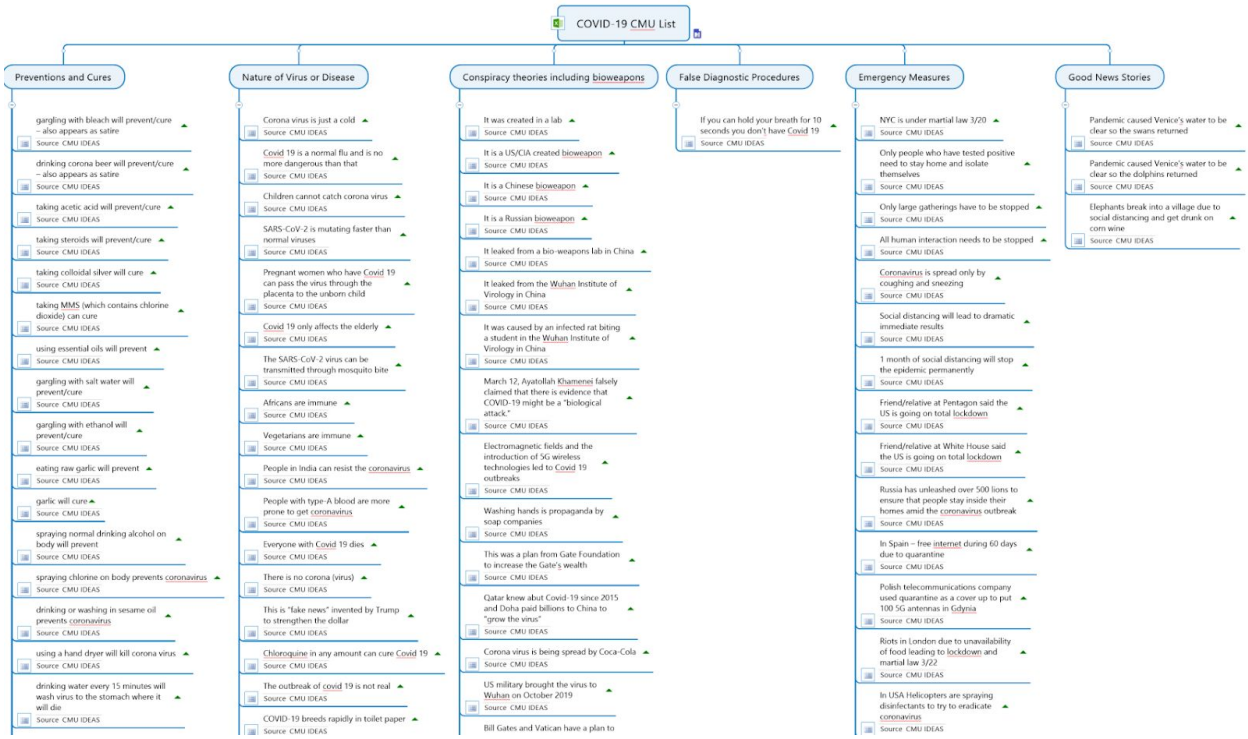
- Tasks (not shown): things that need to be done at each stage. Tasks are things you do, techniques are how you do them.

AMITT is now built into the MISP tool.

Social Media Object Models

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



Mindmap of Covid19 Narratives

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.

Further Reading

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
- 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

AMITT

- AMITT Design Guide
- 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