Building standards for misinfosec

Applying information security principles to misinformation response

Credibility Coalition: Misinfosec Working Group
Six Month Report

August 27, 2019

1 Introduction **5**

1.1 The Misinfosec Working Group 5

1.2 This Report 6

1.3 Glossary 7

2 Online Misinformation as an ecosystem **10**

2.1 Misinformation 10

2.2 The Evolution of Misinformation 11

2.2.1 A Brief History of Misinformation 11

2.2.2 Modern Misinformation 12

2.3 Different views of misinformation 12

2.3.1 Information Operations as Instruments of National Power 13

2.4.2 Social Factors 14

2.5 A Structural View of Misinformation 14

3 Misinformation meets Information Security **18**

3.1 Misinformation Parallels with Information Security 18

3.2 Adapting Information Security Frameworks 19

3.3 Threat Intelligence Sharing Organisations 20

3.3.1 ISACs 20

3.3.2 ISAOs 21

3.3.3 CyberInterpol 22

3.4 Existing Work on Misinfosec 22

4 Describing Misinformation Incidents **23**

4.1 Components of Influence Campaigns 23

4.1.1 Actors and Objectives 23

4.1.2 Message, Delivery and Propagation 23

4.2 Social Network Architecture and Message Contagion 23

4.3 Misinformation Campaigns and Incidents 24

4.3.1 Campaigns: Advanced Persistent Threats 24

4.3.2 Incidents: Building-blocks of Campaigns 24
4.4 Analyzing Misinformation Incidents 25
4.4.1 Designing Incident Templates 26
4.4.2 Incident Gathering 26

5 The Hunt for a Misinformation Framework 28
5.1 Misinformation Work Lacked Frameworks 28
5.2 Mappings between Frameworks 28
5.3 The Cyber Killchain 29
5.4 ATT&CK 30
5.5 Marketing Frameworks 30
5.6 Psyops 31
5.7 Misinformation Frameworks 32
5.8 Other Misinformation Models 32
5.9 Why Information Security? 32

6 Building a Usable Misinformation Framework 34
6.1 Methodology 34
6.2 Keeping Compatibility with Existing Infosec Tools 34
6.2.1 The ATT&CK-based Strawman 34
6.3 Finding Techniques in Misinformation Incidents 35
6.3.1 Technique Analysis 35
6.4 Mapping Techniques to Stages 35
6.5 Procedures: Finding New Forms of Attack 36
6.6 Model Cleanup 36

7 The AMITT Misinformation Framework 37
7.1 What AMITT is and Where to Find it 37
7.2 AMITT Components: Phases, Tactics, Techniques 37
7.3 Phases 38
7.3.1 Planning 38
7.3.2 Preparation 39
7.3.3 Execution 39
7.3.4 Evaluation 39
7.4 Describing an Incident with AMITT 39
7.5 Adding to AMITT 39
7.6 Adapting AMITT 39
7.6.1 The AMITT Codebase 40

8 Misinformation Reporting and Response Protocols 41
8.1 Messaging Standards 41

8.1.1 STIX and TAXII formats 41
8.1.2 Mapping from STIX Infosec to STIX Misinformation Standards 42
8.1.3 MISP Formats 43

9 Use Cases 44
9.1 Red-Team, Blue-Team Exercises 44
9.2 Alerting and Defense 45
9.3 Counterattack (and its Limitations) 46

10 Conclusions and next steps 47
10.1 Challenges 47
10.2 Next Steps for the MisinfosecWG 47
10.2.1 Testing/ refining the framework 47
10.2.2 Blue Team Playbook 47
10.2.3 Support to emerging Response Centers 48
10.3 Thanks 48

11 References 49
11.1 MisinfosecWG Outputs to Date 49
11.1.1 Research Papers and Blog Posts 49
11.1.2 Presentations 49
11.1.2.1 Pending 50
11.2 Referenced Documents 50

Executive Summary

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.

We will describe our first 6 months of work in the document to follow.

1 Introduction

1.1 The Misinfosec Working Group

Members of the Credibility Coalition [0] established the Misinfosec Working Group (MisinfosecWG) in December 2018 in order to build standards for sharing data about misinformation incidents and how to respond to them. After reviewing a number of options to bootstrap our model, we decided to base our new model on existing standards in information security.

Since significant portions of the misinformation threat surface are cognitive or social in nature, misinfosec (*misinformation* plus *infosec*) constitutes not just a new discipline, but an extensive interdisciplinary space ripe for collaboration. As Renee DiResta has rightly pointed out, the fight against misinformation is a “whole of society” problem [73].

In January 2019, MisinfosecWG drafted a Mission Statement for the first six months of work, which is repeated in the panel below.

> *The CredCo Misinfosec Working Group (“wg-misinfosec”) aims to develop a framework for the understanding of organized communications attacks (disinformation, misinformation and network propaganda). Specifically we would like to promote a more formal and rigorous classification of*
>
> - *Types of information-based attacks; and*
> - *Types of defense from information-based attacks*
>
> *Among the operating assumptions of the group will that social and cognitive factors can "scale up and down" within the framework—facilitating some definitional and procedural crossover in both the construction of a framework for understanding these attacks and in their detection. In this sense scales might be formulated as:*
>
> - *ACTIONS: What are the atomic "actions" in propaganda attacks?*
> - *TACTICS: How do actions combine to form larger events, including more complex actions and "attacks"?*
> - *STRATEGY: How do the instances of attacks and actions combine to form "campaigns".*
>
> *The main objectives of the group will be to:*

Define major terms of art at focal points on the scale, with an emphasis on descriptive or procedural rigor; Outline the state-of-the-art "Blue Team" options for defense and counter-attack

Panel 1: MisinfosecWG Mission Statement (December, 2018)

In our first six months, we:

- Collected and analyzed over 63 incidents
- Developed a STIX-inspired format for incident reporting
- Convened in Atlanta to organize TTPs and red team incident planning
- Published our framework proposal; presented to Webconf 2019
- Presented to numerous state, treaty and NGO institutions
- Generated several blog posts and public interest publications
- Created AMITT, a stage-based framework for misinformation reporting and response
- Published AMITT as an open source project on Github

This report describes those first six months of work: what we produced, what we found, and where we expect Misinfosec (both the group and the discipline) to go next. In some respects, we have over-delivered. In others, we still have work to do. Most notably, near-term activity should focus on blue team research and exercises which thoroughly explore the space of potential inoculations and counter-attacks.

We are misinfosec.

Co-Chairs: Sara-Jayne Terp, Christopher R. Walker, John Gray
Treasurer: Danielle Deibler
Working Members: Pablo Breuer, Renee DiResta, Chau Tong, Olya Gurevich, Courtney Crooks, Daniel Black, Tom Taylor, Maggie Engler

Our first output is the AMITT misinformation framework. The latest version of AMITT is held in a Github repository [89].

- https://github.com/misinfosecproject/amitt_framework

1.2 This Report

In this report:

- We describe the problem, the thinking, the methodologies and the work done as part of the emerging discipline of MisinfoSec.

- We Introduce the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) response framework for misinformation
- We propose AMITT as the basis of new misinformation response centres, including ISAOs (Information Sharing and Analysis Organizations) and ISACs (Information Sharing and Analysis Centers).

The intended audience of the report is organizations and individuals responsible for research or response in the misinformation space.

1.3 Glossary

We reference concepts from social science, psychology, information operations, information security, marketing and other overlapping disciplines, many of which have different definitions for the same terms. We use the following definitions for the terms in this text.

Term	**Definition**
Framework	A framework is the basic structure underlying a concept, often used to understand that concept and share information about it. In its nascent form, our framework offers an opportunity to construct an empirical typology upon which we hope to build a much-needed theory of misinformation.
Model	To describe misinformation, we adapt stage-based model theories that are based on the ideas: that elements in systems move through a pattern of distinct stages over time; that these stages can be described based on their distinguishing characteristics and; that successive stages build on earlier stages, integrating their achievements into their assumptions.
Misinformation Incident / Misinformation Campaign	We use *misinformation incident* (and *misinformation campaign*) to refer to the deliberate promotion of false, misleading or mis-attributed information […] We are especially interested in misinformation designed to change beliefs in a large number of people. [92]
Disinformation	Disinformation is false information that is deliberately created or disseminated with the express purpose to cause harm. Producers of disinformation typically have political, financial, psychological or social motivations. [92]
Misinfosec	Misinfosec is an emerging discipline, where misinformation defenders intersect with information security specialists.
Information operation	Information operations, also known as influence operations, include the collection of tactical information about an adversary as well as the dissemination of propaganda in pursuit of a competitive advantage over an opponent. [93]

Information security	Information Security (IS) is designed to protect the *confidentiality*, *integrity* and *availability* of computer system data from those with malicious intentions. This triad has evolved into what is commonly termed the Parkerian hexad, which includes *confidentiality*, *possession* (or *control*), *integrity*, *authenticity*, *availability* and *utility*. [94]
Incident	Incidents are shorter-duration attacks such as "Pizzagate", which could themselves be part of a campaign,
Campaign	We refer to longer, sustained attacks (such Russia's interference in the 2016 US elections) as *campaigns*.
Narrative psychology	Narrative psychology provides additional mechanisms by which to interpret why individuals or groups choose to act in specific contexts. A narrative that grips one audience enough to transport them into action or inaction is very seldom the same narrative that can transport a different audience [95]. Note that this conception of narrative psychology overlaps and aligns with *socio-technical systems* (STS).
Story arc	A misinformation campaign comprises an extended or continuing storyline in episodic storytelling, with each episode following a dramatic arc building on previous episodes. For example, British sanctions following the Skripal poisoning followed the British historical tradition going back the Crimean war of unfairly targeting Russia for aggression.
Artifact	Artifacts are the Images, text, videos, audio assets and websites that adversaries deliver to their target audiences.
Tactic	A tactic is the highest-level, most coarse-grain, description of an actor's behavior.
Technique	Techniques describe how tasks are carried out. Techniques can include relatively benign approaches, such as the use of humour and rhetoric, or more sinister activity such as the forging of documents or the use of false identities [92].
Procedure	A procedure is a lower-level, highly detailed incident description in the context of a technique.
Boom	The concept of *boom* partitions the killchain stages based on whether they occur before or after the detrimental impact of the incident is unleashed on the target. "Left of boom" includes activities in preparation for the attack, such as establishing command and control, cultivation of sock-puppet social media accounts or researching target populations. By contrast, "right of boom" refers to the aftermath. We borrowed these terms from the military, referring to the instant that explosives detonate, as well as the preceding

	and succeeding moments [92]. For our purposes, a boom amounts to cognitive or social impact on the target population.

Panel 2: Major Misinfosec Definitions

2 Online Misinformation as an ecosystem

2.1 Misinformation

Disinformation is commonly defined as dissemination of explicitly false or misleading information; and *misinformation* as the communication of false information without intent to deceive, manipulate or otherwise obtain an outcome [1]. Elsewhere, *misinformation* is used for the deliberate case, both inclusive and exclusive of the accidental case. Other researchers use *malinformation* to describe information that misleads by lacking proper context for interpretation.

While attacks are inherently intentional, the spread of false information by individuals within an attack may not be. At any rate, we will not explicitly use terminology to draw this distinction here, because intentionality of the misinformation propagating individuals isn't pertinent to any ultimate harm. That said, we will pause for a moment to address a common critique that claims the expression *misinformation attack* is incoherent, based on the technical definitions of *disinformation* and *misinformation*. It should be clear from our choice of terminology that we do not entirely agree.

1. *Misinformation* is the commonly-used expression for denoting the superclass in folk language.
2. *Misinformation*, *disinformation* and *malinformation* should be used carefully as technical terms, but worrying too much about their use in derivative terminology is excessively pedantic. Worse too rigid an adherence to traditional naming and approaches could discourage new, cross-disciplinary participation. Often the terminological hand-wringing gives the strong impression that territorial concerns are forefront in the minds of critics.
3. Misinformation attacks often make extensive use of amplification within existing information environments. Since these amplifying populations are often unaware of their informational error, teasing apart mis- and disinformation amounts to a purely semantic exercise with little to gain—modern tactics are simply orthogonal to this dimension.
4. Similarly, truth is often less important than other (social and cognitive) factors in the impact of an attack. Whether the attacker is aware of their "error" becomes similarly unimportant in this context.

We use *misinformation incident* (and *misinformation campaign*) to refer to *the deliberate promotion of false, misleading or mis-attributed information*. In other words, we are talking about the deliberate promotion of mis-, dis- and malinformation. While these incidents occur in many venues (print, radio, etc), we focus on the creation, propagation and consumption of misinformation online. We are especially interested in misinformation designed to change beliefs in a large number of people, or targeted at influential individuals, such as policymakers, activists and journalists.

2.2 The Evolution of Misinformation

2.2.1 A Brief History of Misinformation

Misinformation and psychological operations are as old as war itself. Sun Tzu advocated the use of misinformation. Alexander the Great was known to have directed the manufacture of oversized breast plates for non-existent giants to discourage revolts from conquered territories.

The first information revolution came with the invention of the Gutenberg press. Prior to movable type, creating a written message was a manual process requiring specialized knowledge (literacy) which was both uncommon and laborious. Movable type allowed mass production of messages, but the transmission range was limited by both the ability to physically distribute the model, and the ability to receive (e.g. read) the message. The ability to create content was confined to rulers or large entities (e.g. the church) who had the financial means to produce and transmit their message. The scarcity of messages led to lack of conflicting messages and therefore some level of assurance of authenticity.

The telegraph continued the evolution of information and communication technologies by allowing for the faster transmission of messages that covered a larger geographic area. Although literacy had increased greatly, knowledge of Morse code was still a specialized skill, and telegraph locations were controlled by a few large entities (e.g. American Telephone & Telegraph company) who could simply refuse to transmit or forward a message.

The advent of radio similarly increased the area that could receive a message, but did not fundamentally change who could transmit (mis)information as long range radio transmitters still required significant resources. Radio allowed for anyone with the correct equipment to receive (mis)information in their home. The scarcity of ability to transmit to a wide audience led to a belief in the authenticity of the transmission and this public faith in radio broadcasts led to widespread panic when in 1938, the listening public failed to recognize the War of the World's broadcast as misinformation.

Television was the next major advancement. Now (mis)information could be received without any specialized knowledge, yet the ability to transmit was limited further. Television brought with it the ability to transmit visual misinformation as well as printed or spoken word. If someone wanted to address the whole of the U.S. populace in the 1980's, they needed to be someone who could compel the television stations to carry their message. Certainly the President could do so by threatening the TV station's FCC license to transmit. Radio and television also allowed for limited interaction between those transmitting messages and those who received them. If a member of the audience didn't like a message being transmitted, they could write their station via post. The station could select which letters they would respond to and therefore allow for some amount of censorship.

2.2.2 Modern Misinformation

While misinformation continues in the many traditional venues (print, radio, etc), the creation, propagation and consumption of misinformation online has had an outsized impact that will be the focus of our work. The Internet has significantly accelerated the information evolution started by movable type. Information technologies prior to the Internet enabled for wider dissemination of a message created by a select few, allowing for anyone to instantly transmit a message to a broad audience, without censorship or gate-keeping. For instance Edward Snowden and the Shadow Brokers reached mass audiences, and entertainers Katy Perry or Taylor Swift have more than double the Twitter audience of the President of the United States or the new Prime Minister of Britain, but are constrained by none of the norms of international diplomacy. They each can readily transmit a message to in excess of fifty million recipients.

Recent misinformation campaigns have had a scale and scope previously unseen outside of a declared world war. The use of the cyber domain and social media has reduced dramatically the resources and speed required for misinformation incidents to be created, conducted, evaluated and honed, and is driving the upsurge in misinformation campaigns. Computational propaganda techniques have allowed for the instantaneous transmission of (mis)information to a variety of diverse populations in a highly targeted way on a scale from large subsections of an electorate to a small number of elite administrators or politicians. The ability to provide and gauge feedback to misinformation messaging (e.g. “likes”, forwards, etc.) allows for rapid testing and honing of messaging to achieve the desired effect in near-real time.

Actors behind misinformation incidents include nation-states, institutional actors, grassroots trolls, and financially-motivated freelancers. Common motives include advancement of geopolitical aims, issue-promotion, and financial gain. Governments worldwide are studying misinformation as a form of influence operation or information war [3][4]. In the run-up to the 2016 US election, websites churning out fabricated stories were a cottage industry [5].

2.3 Different views of misinformation

Authors and researchers have described the misinformation problem from a number of different perspectives:

- Information security (e.g. Gordon, Grugq, Rogers) [18][19][20][21]
- Information operations / influence operations (e.g. Kerr, Lin) [22][23][24]
- A form of conflict (e.g. Singer, Gerasimov)
- A social problem
- News source pollution

And each of these perspectives assumes different models and terminologies to address the same issues.

2.3.1 Information Operations as Instruments of National Power

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

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

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

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

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

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

Democracies and autocracies appear to have different vulnerabilities to information threats [72][75][76]. 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.

2.4.2 Social Factors

Social factors are an important consideration for both attackers and defenders in the misinformation space. For example, pathways of communication and attention across social media play a significant role in determining the extent that a given piece of information will spread. These dynamics often include highly-textured interactions between mainstream media, propagandists on legacy platforms, and their social media sock-puppets and supporters, referred to in Benkler [1] as *network propaganda*.

The types of informational networks that serve as a substrate for the propagation of these messages vary significantly. Some networks are *truth-corrective*: they sanction information that diverges from verified reporting. Other networks are *narrative-corrective*: they sanction information that diverges from the consensus narrative of the network. The role of these network-driven attention frameworks on the promotion of marginal framings into the mainstream should not be underestimated. Journalists and policymakers are often directly targeted for harassment or promotion on twitter.

Successful responses to misinformation incidents must be crafted in the context of these varied actors, targets, messages, goals and networks.

2.5 A Structural View of Misinformation

We present here a conceptual decomposition of misinformation into its component pieces, which we have found useful.

We will refer to individual examples of false or misleading content as *misinformation artifacts*. Examples of artifacts would include a tweet with the statement “Barack Obama was born in Kenya”. From an artifact like this, we can discover facts about the attackers (and the threat surface), such as their birther-related messaging goals and the use of particular sock-puppets belonging to a specific person projecting an authentic voice. Each artifact comprises many of a campaign’s dimensions, exposing content, actors, targets and resources to a varying degree from instance to instance.

Misinformation artifacts are combined into or preferably grafted onto narrative frameworks, i.e. stories, that endow them with meaning and emotive content, and typically present a course of action, for example a political donation, a vote or the armed storming of a pizza parlor [6]. In turn multiple narratives may be coherently deployed in a misinformation incident directed at some specific target. In turn again, multiple incidents, directed against potentially multiple targets may form a misinformation campaign (e.g. Russia’s 2016 US election interference and misinformation measures) to achieve a specific goal. Often it is only the artifacts (fake accounts, messages etc) that are immediately visible to incident responders [41][46], while the narratives, incidents and campaigns must be deduced from these, and often only after the fact.

This layering of components to achieve a singular purpose may be viewed as a pyramid, pictured in *Figure 1* below. Each layer of the pyramid, in addition to being about misinformation, is also about the actions taken by the attacker to achieve it, and the actions the attacker desired to elicit in the target.

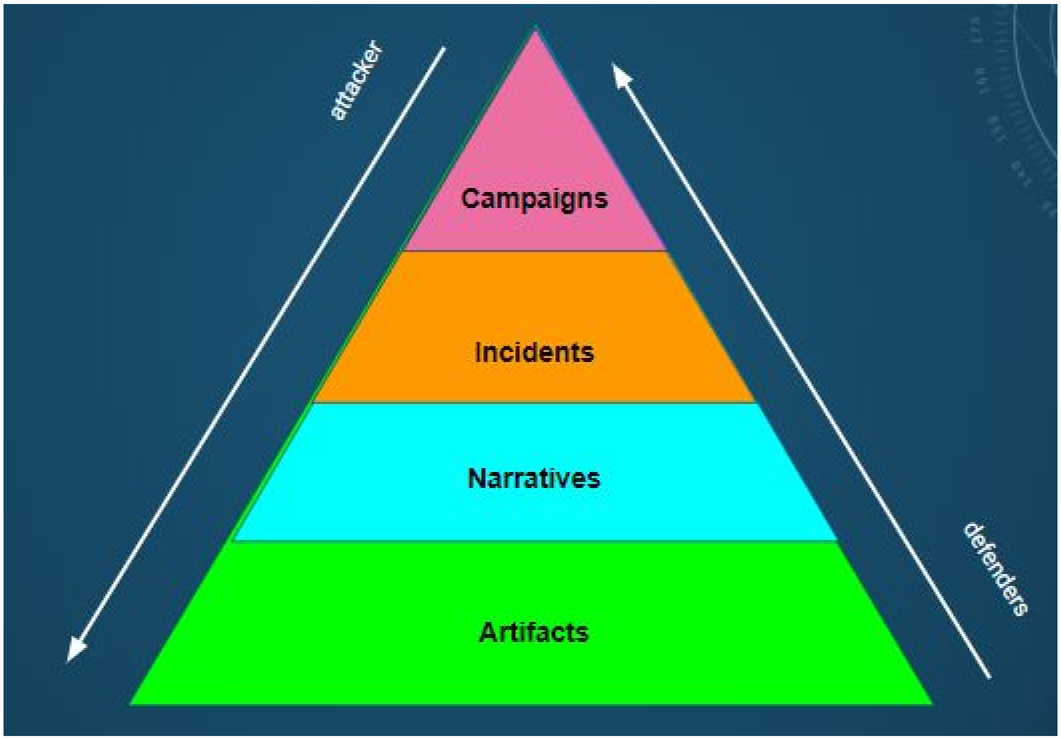


Figure 1: misinformation pyramid

Our model assumes three main groups: the promoters of misinformation (*attackers*), the targets of the misinformation (*populations or individuals*), and the people hopefully trying to counter them (*defenders*). Attackers create *incidents* (e.g. Macrongate), which often form part of longer-term *campaigns* (e.g. destabilize French politics). Human communication generally takes place at the level of 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.

The attacker sees the whole of the pyramid from the top down, but the defender must reconstruct it from the bottom up. Unless they're lucky enough to have good insider information or intelligence, defenders must work back from artifacts to understand incidents and campaigns. Among other things, this informational asymmetry means that most current misinformation work is at the artifact level. Recent work on narratives has endeavored to expand the defenders' visibility upward [46][96], while work by this group and others have attempted to describe specific routes to the summit via red team exercises and incident analysis. But note that acceptance of the informational (or narrative) content of a misinformation attack may not be its objective. The specific contents of belief are often incidental to the attacker's desired goal, which could better be described as creating discord and confusion.

3 Misinformation meets Information Security

"Somebody's always had control over information, and others have always tried to steal it. Read Machiavelli. As technology changes, sneakiness finds new expressions."
*— **Clifford Stoll**, The Cuckoo's Egg*

Misinfosec represents the intersection of "misinformation" and "information security." We adapted the term from cybersecurity strategies and use it to refer to the fight against the spread of false information. Whether attackers are trying to spread malware or disinformation, they have to study the people they're targeting. After creating convincing artifacts such as images, text, and websites, attackers must present those artifacts to their target audiences. Like infosec, we position the defender as someone trying to work out the steps an attacker would have gone through, working from the artifacts up to the campaigns.

3.1 Misinformation Parallels with Information Security

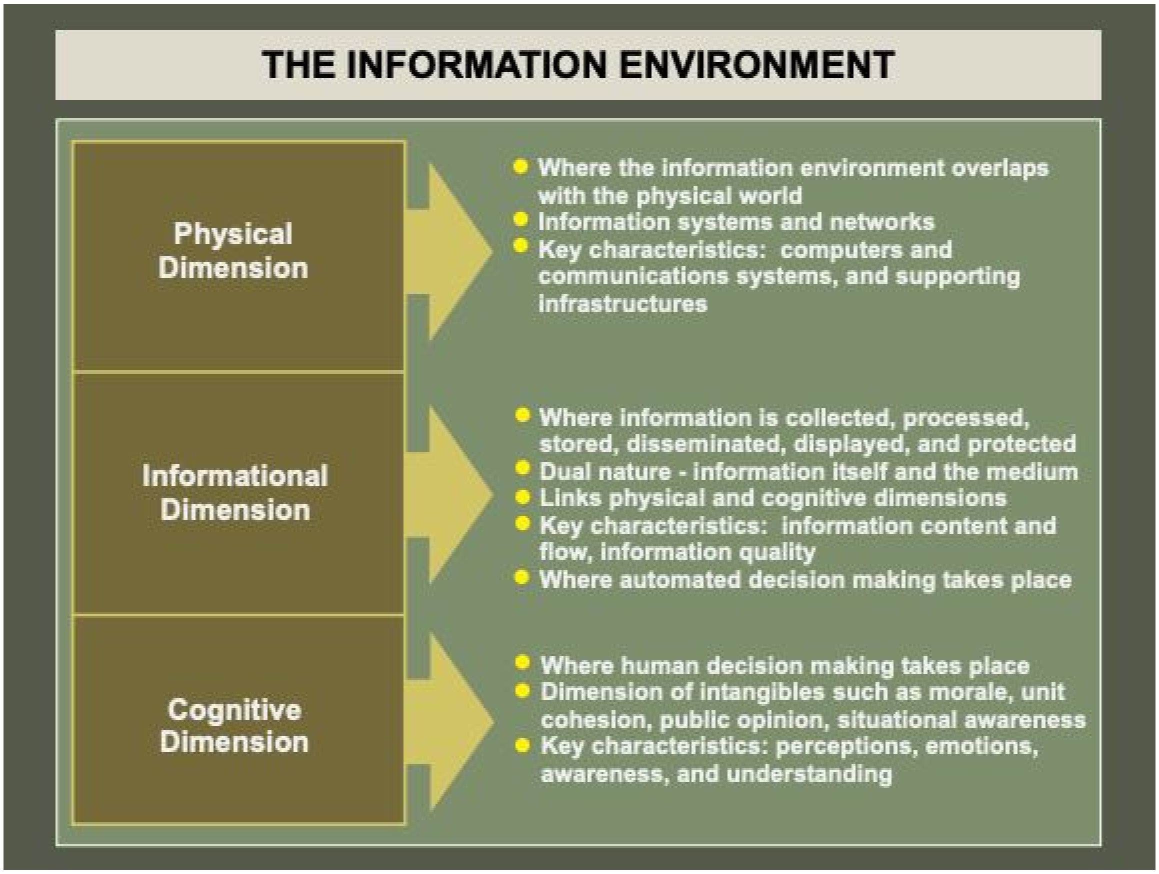
The pursuit of networked information security all started one day in 1986 with a $0.75 accounting error. Clifford Stoll traced the error to an unauthorized user who had apparently used 9 seconds of computer time and not paid for it. What followed was a chain of events revealing the first documented case of a state-sponsored cybercrime. In 2013 Mitre Corp. started the ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge) framework as a way to categorize common adversary behavior for Cyber Security research. It took 27 years to create a framework that deals with the bad-actors hacking our most valuable networks and stealing our most vital secrets [36][37][48].

Fast forward to 2016 and consider the results of both Brexit and the US Election. A study of the antecedents to these events lead us to the realization that there's something off kilter with our information landscape. The old cold war political warriors, their hackles rising, would recognize an intimately familiar situation now deep in an uncanny valley of networked tooling. The usual useful idiots and fifth columnists—now augmented by automated bots, cyborgs and human trolls—are busily engineering public opinion, stoking up outrage, sowing doubt and chipping away at trust in our institutions. And now it's our brains that are being hacked.

We do not need, nor can we afford, to wait 27 years for the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) framework to go into use. Misinformation security is born as a recognized discipline that is related to and borrows lessons learned in information security but is distinct from it in terms of domain of operation and techniques. Misinfosec can play a significant role in moving beyond admiration (papers and talks) of the online misinformation problem and into a standards-based community that is taking collective action to mitigate the intentional harms being perpetrated by adversarial state actors and their proxies.

3.2 Adapting Information Security Frameworks

Information security encompasses offensive and defensive computer network operations, electronic warfare, psychological operations, military deception, and operational security [33]. Information security is a robust field with well-understood principles and best practices, covering physical, informational and cognitive dimensions of the information environment.



THE INFORMATION ENVIRONMENT

Physical Dimension

- Where the information environment overlaps with the physical world
- Information systems and networks
- Key characteristics: computers and communications systems, and supporting infrastructures

Informational Dimension

- Where information is collected, processed, stored, disseminated, displayed, and protected
- Dual nature - information itself and the medium
- Links physical and cognitive dimensions
- Key characteristics: information content and flow, information quality
- Where automated decision making takes place

Cognitive Dimension

- Where human decision making takes place
- Dimension of intangibles such as morale, unit cohesion, public opinion, situational awareness
- Key characteristics: perceptions, emotions, awareness, and understanding

Figure 2: Dimensions of The Information Environment

Alerting systems exist on top of frameworks and standards describing information attacks like DDOS, viruses, and unwanted internet traffic like spam etc. These systems offer a good place to start with misinformation [10].

We researched potential fits between misinformation and several common information security frameworks: strategic-level models like the *SANS sliding scale* (architecture, passive defense, active defense, intelligence, offense), *Gartner cycle* (prevent - detect - respond - predict), *NIST framework* (detect - protect - identify - recover - respond) and *Cyber Attack Lifecycle*, and operational-level models like the *ATT&CK matrix* and *SANS top 20* [9].

The Cyber Attack Lifecycle [34] basically maps to our campaign descriptions (*reconnaissance*, *weaponization*, *installation*, *exploitation*, *command-and-control*, and *actions on the objective*). We could also use this lifecycle to link goals and intent (e.g. the "four Ds" of propaganda: Dismiss, Distort, Distract, Dismay [35]).

The MITRE ATT&CK Matrix [36][37][48] covers the last three stages of the cyber attack lifecycle, and lists tactic phases (initial access, execution, persistence, privilege escalation, defense evasion, credential access, discovery, lateral movement, collection, exfiltration, command and control) with a set of techniques that an adversary could use in each phase (e.g. *Spearfishing Attachment* is a type of *Initial Access* tactic). The ATT&CK database provides examples, detection and mitigation for each technique, along with extensive references.

The ATT&CK Tactics, Techniques and Procedures (TTPs) describe *patterns of activities or methods associated with a specific threat actor or group of threat actors*. *Tactics* are the top-level steps that an attacker typically takes (e.g. "amplify message"); *techniques* are the different ways those steps can be done (e.g. "repeat message using bots"); *procedures* are the sequences of actions in an attack.

ATT&CK TTPs were created by taxonomizing existing information security threat reports and analyses. MisinfosecWG used a similar process of assembling and grouping TTPs in support of blue team efforts in Misinformation Security (Misinfosec).

3.3 Threat Intelligence Sharing Organisations

3.3.1 ISACs

Presidential Policy Directive 21 (PPD-21 [61]) is the critical infrastructure protection and resilience directive. PPD-21 identifies sixteen critical infrastructure sectors. As part of defending these critical infrastructure sectors, the U.S. Department of Homeland Security helped establish an Information Sharing and Analysis Center (ISAC) for each. The ISACs are non-profit organizations that provide a central hub for gathering and sharing of cyber threat information to those sectors.

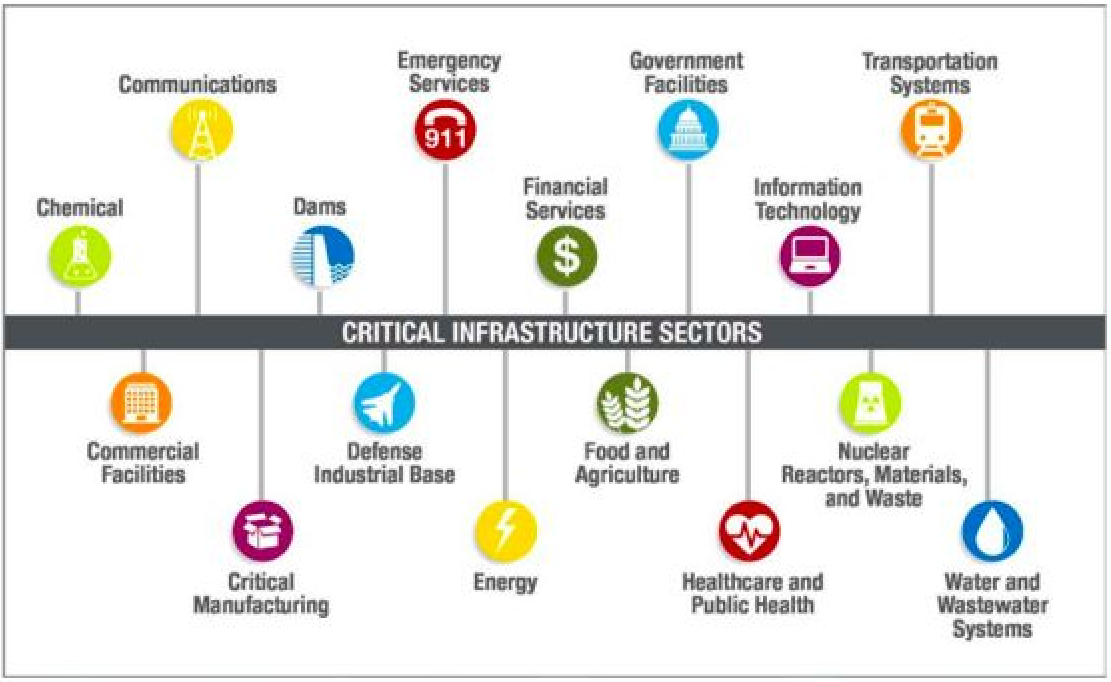


Figure 3: Critical Infrastructure Protected by ISACs [61]

Some things that might be considered social and political critical infrastructure: confidence in the rule of law; confidence in the integrity of elections; the separation of powers; the political control of the military. Maybe these should someday have ISACs of their own.

3.3.2 ISAOs

Any sector that is outside of those sixteen critical infrastructure sectors can establish a similar organization called an ISAO (Information Sharing and Analysis Organization). Functionally, ISAOs and ISACs are the same; they differ only in whether they address threats to critical infrastructure or non-critical infrastructure.

Currently, ISACs track only cyber threats relevant to their respective sectors. Misinformation attacks against these critical sectors can have diplomatic and economic consequences and can even lead to military action. Defending critical sectors against misinformation attacks is of paramount importance to the security of the nation. While social media is not identified as a critical sector, and therefore doesn't qualify for an ISAC, a misinformation ISAO could and should feed indications and warnings into ISACs. The establishment of a misinformation ISAO could enable ISACs to be aware of misinformation threats to critical infrastructure sectors [65][66][67]. However, we argue that perhaps the integrity of our shared informational environment should be categorized as critical infrastructure. Similarly, major components of our social infrastructure like confidence in the rule of law or election integrity must be treated as critical infrastructure.

3.3.3 CyberInterpol

Interpol describes their Cyber Fusion Centre as “an operation that brings together cyber experts from law enforcement and industry to gather and analyze all available information on criminal activities in cyberspace and provide countries with coherent, actionable intelligence.” The Centre publishes reports to alert countries to new, imminent or evolving cyber threats; these include malware, phishing, compromised government websites, social engineering fraud and more. In 2017, they provided 183 reports to police in nearly 70 member countries worldwide, while in the first half of 2018 alone, we disseminated 187 reports to 138 countries [81][82].

3.4 Existing Work on Misinfosec

Our work grew out of earlier work describing red team and blue team misinformation tactics [8] and characterizing misinformation as an information security problem that infosec frameworks and principles could be applied to [9][10][11].

Information security teams increasingly treat misinformation as a subject of concern. FireEye helped characterize the Iranian IUVM disinformation network [12], tracked disinformation typosquats [13] and analyzed traffic in the 2018 US midterm elections [14]; ThreatConnect [15] tracked online infrastructure behind Russian misinformation campaigns, and Synack [16] described how misinformation could be used as part of an information security attack.

Infosec experts have also discussed expanding definitions of information security to include misinformation. Landau [17] argues that the NSPD-54 definition of cybersecurity should be extended to include information operations (e.g. misinformation), and raises the issue of misinformation users adapting their tools and techniques as detection improves. Rogers [18][19] frames misinformation as an information integrity problem, citing the infosec concept of maintaining CIA (*confidentiality*, *integrity*, *availability*), and suggests applying infosec practices such as threat modelling. Brockman and Grugq describe parallels between misinformation and information security [20][21]. Lin and Kerr [22][23][24] examine cyber-enabled information warfare as a conflict form where the USA is weak, examine the environment, operations and characteristics of this space, and call for new tactics and responses in it.

4 Describing Misinformation Incidents

4.1 Components of Influence Campaigns

Benkler et al [1] provide a number of useful distinctions. First, they suggest viewing categories of online information threats in terms of a few scalar dimensions, such as *centralized* versus *decentralized*, *political* versus *commercial* and *technological* versus *institutional*. These dimensions suggest we focus on factors such as the *actors*, *objectives*, and *delivery mechanisms* in describing the terrain.

4.1.1 Actors and Objectives

Influence operations are undertaken by an *attacker*, directed at a *target* and sometimes amplified by *carriers*. These actors can all be individuals, populations or institutions. But ultimately the targets are people, or aggregates of people. The objective is typically tied up in the psychology of the actors, especially the targets and carriers. The goal is to change the beliefs and behavior of individual people, often at scale, via manipulation: *directly influencing beliefs, attitudes, or preferences of a target population in ways that are not normatively appropriate in context.* [1] The cognitive objectives of the operations often include widespread factual misunderstanding or confusion. This can vary from *gaslighting* and *disorientation* to *distraction, priming* and *agenda setting*.

4.1.2 Message, Delivery and Propagation

The tactical objective of an attack is typically cognitive or social. The payload is information. Misinformation is semantically misleading, contextually misleading, misattributed, or factually incorrect. The information is transferred via communication events, which themselves can be varied and subtle; some are harder to address than others. The information can range from propaganda to bullshit. *The bullshit artist "does not care whether the things he says describe reality correctly. He just picks them out, or makes them up, to suit his purpose."* [32]

4.2 Social Network Architecture and Message Contagion

The social factors at play involve pathways of communication and attention. For the most part, this amounts to social media. But the interaction between "mainstream" propagandists and their social media sock-puppets and supporters is highly textured. Benkler [1] refer to this interaction as *network propaganda*, noting the familiarity effect with this approach. Simply repeating a message enough in the right forums is adequate to cement its content in the minds of many listeners.

Moreover, the types of informational networks that serve as a substrate for the propagation of these messages vary significantly. Some networks are *truth-corrective*: they sanction information that diverges from verified reporting. Other networks are *narrative-corrective*: they sanction information that diverges from the consensus narrative of the network. Networks for journalists on Twitter tend to be truth-corrective, whereas QAnon tends to be narrative corrective. Most networks exhibit a blend of these pressures: sometimes responsive to accuracy demands; other times responsive to more normative concerns.

The role of these network-driven attention frameworks on the promotion of marginal framings into the mainstream should not be underestimated. Benkler also describes an attention backbone which promotes stories from the periphery of the network and *propaganda feedback loops* which are *pathological network dynamics in which (mostly attentional) sanctions are imposed for breaking with the received narrative preferred by the target population.* Meanwhile journalists online are targeted for harassment, priming and other forms of influence.

In short, the objectives are cognitive and the vectors of delivery are social. Successful responses to attacks must be crafted in the context of these varied actors, targets, messages, goals and networks.

4.3 Misinformation Campaigns and Incidents

4.3.1 Campaigns: Advanced Persistent Threats

In information security, an Advanced Persistent Threat (APT) is an attack or an attacker operating over a long period of time. APTs are usually (but not always) backed by nation-states. In misinformation, APTs usually run long-duration campaigns. Watts [70] updates this term to *Advanced Persistent Manipulator*. The canonical nation-state misinformation campaign is the 2015-2017 Russian troll farm work on the 2016 US presidential elections. Jamieson [38, p. 75] describes these operations in detail. The objectives are numerous: *an amplifying effect; an agenda-setting effect; a normative effect; target identification; a mobilizing effect; a two-step flow effect; weighting, contagion and spiral of silence effects;* and *a familiarity effect.*

Benkler [1] looked at the online spread of prominent political stories before and after the 2016 US presidential elections. While the authors discuss aspects of the Russian campaign, their primary focus is the online media ecosystem itself. First, the online political environment in the United States is polarized, but the “filter bubbles” are best characterized as 1) the Fox News bubble and 2) everyone else. Second, the corrective sanctions at play in the two environments are highly asymmetric. The latter bubble penalizes for straying from the truth while the former penalizes for straying from the accepted narrative. Interventions against misinfo attacks in these two environments could be very different.

4.3.2 Incidents: Building-blocks of Campaigns

Campaigns are often built from smaller building blocks. We will refer to those as *incidents*. One example is the 2014 Columbian Chemicals incident, which we’ve listed as:

Summary	Early Russian (IRA) “fake news” stories. Completely fabricated; very short lifespan.
Actor	probably IRA (source: recordedfuture)

Timeframe	Sept 11 2014 (1 day)
Presumed goals	test deployment
artifacts	text messages, images, video
Method	1. Create messages. e.g. “A powerful explosion heard from miles away happened at a chemical plant in Centerville, Louisiana #ColumbianChemicals” 2. Post messages from fake twitter accounts; include handles of local and global influencers (journalists, media, politicians, e.g. @senjeffmerkley) 3. Amplify, by repeating messages on twitter via fake twitter accounts
Result	limited traction
Counters	None seen. Fake stories were debunked very quickly.
Related attacks	These were all well-produced fake news stories, promoted on Twitter to influencers through a single dominant hashtag: *#BPoilspillTsunami* *#shockingmurderinatlanta* *#PhosphorusDisaster* *#EbolaInAtlanta*

Panel 3: Sample Incident Report

4.4 Analyzing Misinformation Incidents

MisinfosecWG is analyzing known misinformation campaigns. A *campaign* is online manipulation designed to influence the beliefs of a large number of people. It typically consists of several attacks or incidents, aligned toward a specific goal.

In 2017, at least 18 countries used misinformation tactics in elections [25]. Most employed groups of “opinion shapers” to manipulate domestic elections; some, including Russia and Iran, used these tactics to manipulate popular beliefs in other countries [26]. The more well-known campaigns include Russian interference and influence on Brexit [27], the 2017 French Presidential election, and the 2016 election of Donald Trump; attacks on the Parkland teenagers; promotion of Jade Helm conspiracy theories; and various influence operations around the Black Lives Matter movement [28]. Private influence operators [30] manipulate beliefs; grassroots 'trolls' and marketers use misinformation campaigns to push agendas or make money, usually from online advertising. And there are countless less well-known international operations [29][70][76].

With few exceptions [31], most of the response thus far has been akin to whack-a-mole. The extent of the threat and the range of possible actions are simply not understood well enough to formulate counter-moves, whether tactical or strategic.

In our analysis, we look at the actors and their presumed goals and timeframes. We also look in detail at the methods used in each attack; look at the counters used against them; and list related attacks, as a first pass at creating attack types that can be grouped and discussed together.

4.4.1 Designing Incident Templates

Our stage-oriented framework is based on an analysis and cataloguing of various known misinformation incidents in the recent past. We used a STIX-based (Structured Threat Information eXpression) template for misinformation incidents, by anticipating framework needs and inspecting our existing collection of incidents and campaigns [49]. The template fields are shown in Panel 4 below.

Name	**Description**
Title	Descriptive title (also used as the filename)
Summary	The TL;DR about this incident or campaign
Actor	Who, if anyone, was suggested as the source of this incident - e.g. Russia, IRA, far right group, individual etc
Timeframe	How long did this incident last, e.g. a year, a few months, a day
Date	Rough date this incident started, or happened if "started" is hard to determine
Presumed goals	What you think the actor was trying to do with this incident
Method	Techniques used
Counters	Actions taken against this incident
Related attacks	Other incidents related to or very similar to this one
References	References to articles, URLs etc. used to complete this form

Panel 4: Incident template fields

4.4.2 Incident Gathering

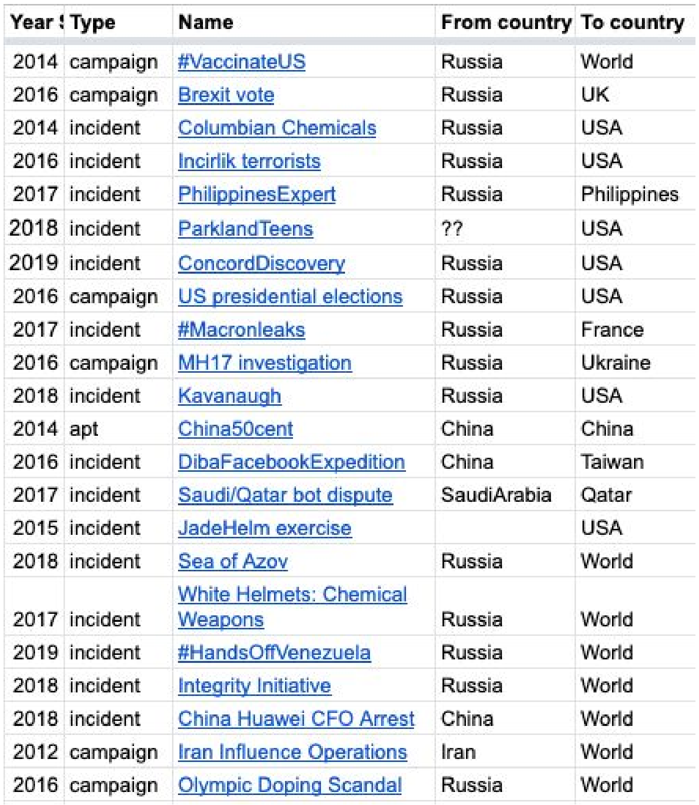
Using our STIX-inspired format, we documented over 63 incidents, of which we selected 22 to serve as the basis of the AMITT framework. We searched open-source accounts and datasets for misinformation incidents and campaigns between 2012 and 2018, screening for less-common incident types, and looking carefully for incidents that did not appear to have attribution to Russia.

Some of the sources used were:

- Academic research (e.g. [57])
- Policy organisations (e.g. [58])
- Other misinformation researchers (e.g. [59])
- Other misinfosec researchers (e.g. [60])

In terms of naming conventions, it became apparent that one term for a coordinated misinformation attack was not adequate, and chose *campaign* for longer attacks, *incident* for shorter-duration attacks, which could themselves be part of a campaign, and *APT* (from the infosec term *Advanced Persistent Threat*) for an attacker that appeared to be engaged in continuous campagins. Like us, the Oxford Internet Institute (OII) calls the Russian work on US elections a "campaign" [57].

From these open source resources, we created an initial collection of incidents, campaigns and APTs, completing a template for each.



Year S	Type	Name	From country	To country
2014	campaign	#VaccinateUS	Russia	World
2016	campaign	Brexit vote	Russia	UK
2014	incident	Columbian Chemicals	Russia	USA
2016	incident	Incirlik terrorists	Russia	USA
2017	incident	PhilippinesExpert	Russia	Philippines
2018	incident	ParklandTeens	??	USA
2019	incident	ConcordDiscovery	Russia	USA
2016	campaign	US presidential elections	Russia	USA
2017	incident	#Macronleaks	Russia	France
2016	campaign	MH17 investigation	Russia	Ukraine
2018	incident	Kavanaugh	Russia	USA
2014	apt	China50cent	China	China
2016	incident	DibaFacebookExpedition	China	Taiwan
2017	incident	Saudi/Qatar bot dispute	SaudiArabia	Qatar
2015	incident	JadeHelm exercise		USA
2018	incident	Sea of Azov	Russia	World
2017	incident	White Helmets: Chemical Weapons	Russia	World
2019	incident	#HandsOffVenezuela	Russia	World
2018	incident	Integrity Initiative	Russia	World
2018	incident	China Huawei CFO Arrest	China	World
2012	campaign	Iran Influence Operations	Iran	World
2016	campaign	Olympic Doping Scandal	Russia	World

Figure 4: Incidents selected for technique analysis

The problem of misinformation is a global one; not just Russia targeting the United States. While admittedly skewed towards incidents with Russia as the aggressor, our exercise has given us a collection of example techniques and artifacts that we can organize into a descriptive framework. Twenty-two of these incidents were selected for further analysis (*Figure 4*).

5 The Hunt for a Misinformation Framework

5.1 Misinformation Work Lacked Frameworks

Analysts and engineers have been creatively adapting their own techniques to a range of issues that they perceive to be essential to confronting the larger problem of online misinformation.

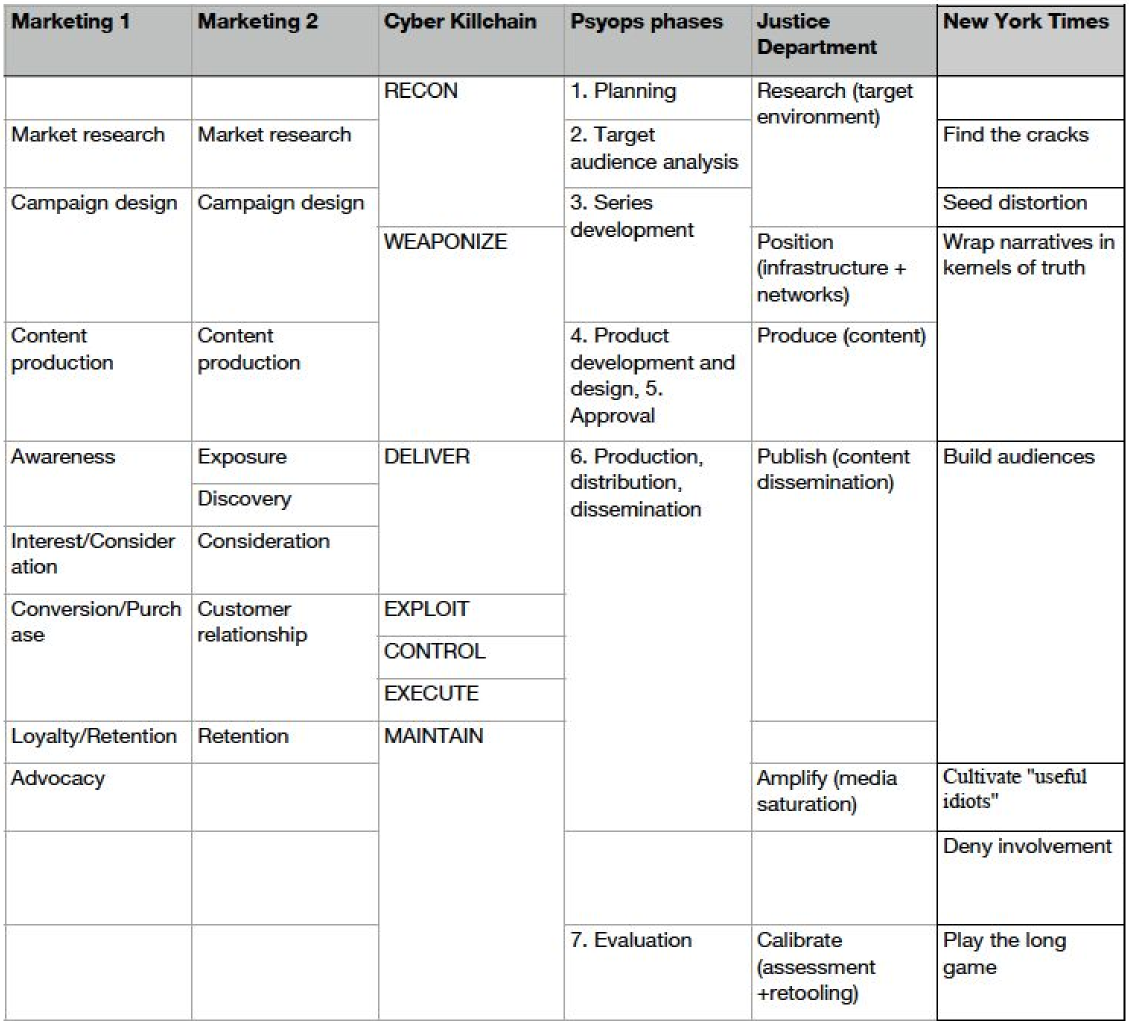
- Which sites promote false stories?
- How can you detect a false story?
- How can you detect a false statement or fake recording?
- How can you detect a doctored photograph or video?
- How can you detect a sock-puppet or a troll?

Some platforms have processes in place for detecting and mitigating nefarious user activity. And independent technical approaches have begun to sprout up as well [5][6][7]. These techniques and toolkits will no doubt have some application in the broader domain of combating misinformation. But tools in themselves have no values and serve no inherent masters.

Misinformation operates within a complicated socio-technical ecosystem, so the approach must be multidisciplinary. Just as cyber security professionals must keep up with constantly evolving techniques and strategies for exploiting private information available within cyberspace, misinformation professionals should expect the same. Further, policy makers must be enabled to develop and enforce data privacy policies that help protect their public from nefarious online activity and privacy breaches. The application of technical and operational techniques can only be productive in the context of thoughtfully designed tactics, grounded in well-defined strategy. Worse, absent strategic and tactical goal-setting we cannot be certain that these techniques are either necessary or sufficient to the amelioration of the threat presented by online influence operations. Without a framework, we are operating in the dark.

5.2 Mappings between Frameworks

Creating a new framework required analysing the current stage-based models and determining the different features, work flows, and overall processes that best fit the MisinfoSec requirements. We define a *framework* as a basic structure underlying a concept, often used to understand that concept and share information about it. Before creating AMITT, we researched existing misinformation frameworks, but found nothing suitable for the variety of misinformation incidents we needed to document. To sanity check our process, we mapped major frameworks to each other (*Figure 5*) and compared AMITT version 1.0 against each of them, looking for stages and/or phases we might have missed.



Marketing 1	Marketing 2	Cyber Killchain	Psyops phases	Justice Department	New York Times
		RECON	1. Planning	Research (target environment)	
Market research	Market research		2. Target audience analysis		Find the cracks
Campaign design	Campaign design		3. Series development		Seed distortion
		WEAPONIZE		Position (infrastructure + networks)	Wrap narratives in kernels of truth
Content production	Content production		4. Product development and design, 5. Approval	Produce (content)	
Awareness	Exposure	DELIVER	6. Production, distribution, dissemination	Publish (content dissemination)	Build audiences
	Discovery				
Interest/Consideration	Consideration				
Conversion/Purchase	Customer relationship	EXPLOIT			
		CONTROL			
		EXECUTE			
Loyalty/Retention	Retention	MAINTAIN			
Advocacy				Amplify (media saturation)	Cultivate "useful idiots"
					Deny involvement
			7. Evaluation	Calibrate (assessment +retooling)	Play the long game

Figure 5: Mapping between potential frameworks

5.3 The Cyber Killchain

AMITT describes a killchain, and was based originally on the ATT&CK framework, itself an expansion of the Cyber Killchain model [47]. The term killchain was originally used as a military term to define the necessary steps required to successfully conduct an attack, where a killchain is a linear structure in which each link in the chain must be executed and any “link” broken results in an attack failure. A killchain doesn’t describe how a task is to be accomplished, only that it must be accomplished.

In our case, AMITT defines not only the killchain required to conduct misinformation attacks, but also defines tactics, techniques, and procedures (TTPs) which describe how the corresponding link can be accomplished. By grouping these TTPs into the ways and means of completing a task in the kill chain, we can achieve a greater understanding of an attack and more effectively analyze the killchain components. Note that the establishment of a framework allows for better analysis both by those defending against a misinformation attack and by those designing a misinformation campaign.

5.4 ATT&CK

We designed AMITT to be familiar to people working with MITRE's ATT&CK framework [48], and to interact with many of its components (e.g. we adopted the same STIX/TAXII data formats [49] used by ATT&CK etc).

The ATT&CK framework is used by the infosec community to share information about incidents. ATT&CK and the Cyber Killchain divide an incident into stages such as *recon* or *exfiltration*. The ATT&CK framework adds more detail to the last three stages of the Cyber Killchain: these stages are known as "right of boom" (this phrase originally referred to the chain of events after an explosion or attack), as opposed to the four "left-of-boom" stages which happen before the bad actors gain control of a network and start damaging it; a period when you still have time to prepare and avert a crisis.

ATT&CK is widely used by infosec guardians, including existing ISACs and ISAOs (information security information sharing bodies). It defines cyber attacks against machines and networks, but fails to address cognitive attacks such as misinformation. As modern misinformation attacks are conducted via the cyber domain using resources which may have been co-opted via cyber attack, AMITT extends ATT&CK by adding framework components to address these cognitive attacks.

5.5 Marketing Frameworks

Marketing funnels are about the journey or a marketing campaign to the end consumer—the person who watches an online video, sees a marketing image online etc. and is ideally persuaded to change their view of or buys something related to a brand. This is a key consideration when listing stages: whose point of view is this? Do we understand an incident from the point of view of the people targeted by it (which is what marketing funnels do), the point of view of the people delivering it (most cyber frameworks), or the people defending against it? We suggest that the correct point of view for misinformation defense is that of the creator/attacker, because attackers go through a set of stages, all of which are essentially invisible to a defender, yet each of these stages can potentially be disrupted.

Furthermore, the objectives of the attacker are not always directly encoded into the content of the artifacts. While disrupting the message is often the result of a successful counter-tactic, the message content is often a means, not the end. A detailed analysis of operational goals gives defenders a wider range of disruptive options, each tailored more effectively to the threats as designed.

Additionally, marketing funnels are "right-of-boom" in that they begin their analysis at the point where an audience is exposed to an idea or narrative and becomes aware of it. This is described as the customer journey, which is a changing mental state, moving from seeing something to taking an interest, to building a relationship with a brand/idea/ideology and subsequently advocating it to others. This same dynamic plays out in online misinformation and radicalisation (e.g. QAnon effects), with different hierarchies of effects that might still contain the *attraction*, *trust*, and *advocacy* phases.

But we can borrow selectively from the marketing funnel and map these stages across to the Cyber Killchain by adding in *marketing*, *planning*, and *production* stages (market research, campaign design, content production) and seeing how well they align with attackers' game plans. With these adjustments—shifts in the direction of infosec models—we can begin planning how to disrupt and deny these left-of-boom activities.

When considering the *advocacy* phase, we see this fitting the *amplification* and *useful idiot* stages. This is new thinking in relation to other misinformation models, and modeling how an "infected" node in the system isn't just repeating a message but might be or become a command node too is something to consider in more detail.

Additionally, misinformation operations often promote more weighty cognitive changes than the decision of which brand to buy for likely pre-existing purchase decision. Marketing strategies may not be appropriate for such highly personal or ideological messaging, especially where short-term effects are desired [97].

5.6 Psyops

Developing the misinformation framework also requires adopting and acknowledging the role of psyops as this point of view centers on the campaign producer. The producer controls every stage, including a step-by-step list of actions from the start through to a completed operation. Hierarchy-aware roles such as getting sign-offs and permissions can also be included.

When looking left-of-boom, psyops (Psychological Operations) models map closely to the marketing funnel, with the addition of a *planning* stage. While right-of-boom glosses over all the end-consumer-specific considerations, in a process flow defined by "production, distribution, dissemination.", this does add a potentially useful evaluation stage. One of the strengths of working at scale online is the ability to hypothesis test (eg. A/B test [50]) and adapt quickly at all stages of a campaign. Additionally, when running a set of incidents, after-action reviews can be invaluable in learning and adjusting the higher-level tactics such as adjusting the list of stages, the target platforms, or determining the most effective narrative styles and assets.

And psyops literature is packed with useful distinctions.

Defensive propaganda is designed to maintain an accepted and operating form of social or other public action; but *offensive propaganda* is designed to interrupt social action not desired by the propagandist or to predispose targets to social action which he desires. *Conversionary propaganda* is designed to change the emotional or practical allegiance of individuals from one group to another; whereas *divisive propaganda* is designed to split apart the component subgroups of a population; and counterpropaganda is designed to refute a specific point or theme of enemy propaganda [97]. Similarly, psyops literature is supportive of our distinction between campaigns and incidents and draws and line between strategic and tactical propaganda.

Contemporary conversations about misinformation countermeasures often focus on *tactical counterpropaganda*, but recent studies suggest that *strategic defensive propaganda* is broadly understood to be more effective.

5.7 Misinformation Frameworks

The Department of Justices model (Malign Foreign Influence Campaign Cycle [51]) clearly presents what each stage looks like from both the *attacker* and *defender* points of view. This model is a solid description of early Internet Research Agency (IRA) incidents, yet is arguably too passive for some of the later ones. When we say “passive”, we mean this model works for “creating and amplifying a narrative”, but we’re fitting something like “create a set of fake groups and make them fight each other”, which takes on a more active and more command & control like presence. Models like this work well for some but not all of the misinformation incidents that we have seen (or expect to see).

When we formulate incident descriptions and map them to the stages, we start asking how our adversaries are exercising functions such as command & control.

5.8 Other Misinformation Models

Frameworks that aren’t shown on the mapping include Watts [52], diResta [53] and Decker’s [54] models, each of which are at a different level of detail to the Cyber Kill Chain.

Decker’s models look at the groups involved in different stages of misinformation, and the activities of each of those groups. These models focus on misinformation campaigns as a series of handoffs between groups: from the originators of content, to command and control signals (via Gab, Telegram. etc.) for signal receivers to post the content to social media platforms, then amplify its messages with social media messages that eventually get picked up by professional media. This has too many groups to fit neatly onto a marketing model, and appears to be on a different axis to psyops and DoJ models, but still seems important.

As a further axis—the stage models we’ve discussed above are all tactical; the steps that (typically) an attacker would go through in a misinformation incident. There are also strategies to consider, including Ben Nimmo’s “four Ds” (Distort, Distract, Dismay, Dismiss—commonly-used IRA strategies [55]), echoed in Clint Watt’s online manipulation generations. In infosec modelling, this would get us into a Courses of Action Matrix [56], as seen in diResta’s model.

5.9 Why Information Security?

We considered adapting frameworks from several fields. This included advertising frameworks, lean enterprise frameworks and information security (infosec) frameworks. We chose the infosec framework because of the close fit between infosec attacks on individual and networks of machines and misinformation attacks on individuals and networks of humans.

The most frequently mentioned alternative is advertising. Tactics and techniques will inevitably draw from numerous fields, advertising prominent among them. For example, advertising test cases might be able to tell us something about the psychosocial motivators that are most effective in advertising. In other words, what techniques are effective in getting people interested or invested in the "product" of the advertisement? Understanding these motivators could be helpful in understanding what captures people's attention in misinformation campaigns.

"These decisions cannot be compared with the choice of a toothpaste, a deodorant, or a cigarette. Advertising succeeds in peacetime because it does not matter; the choice which the consumer makes is of slight importance to himself, even though it is of importance to the seller of the product. A Dromedary cigarette and an Old Coin cigarette are both cigarettes; the man is going to smoke one anyhow." [97]

But while advertising may have much to offer at the tactical level, we ultimately decided that it did not offer an adequate fit to our problems as a framework. Most notably, advertising uses a subset of the attack techniques that we care about, but does not typically convince an audience to do something against their own interest nor can it legally fabricate false facts. Furthermore, a contrarian might note that tactics from the advertising domain would more appropriately be drawn from the world of anti-corporate advocacy. Efforts to ameliorate ad placement and saturation would have quite a bit to say about which responses are most effective, whether regulatory or otherwise.

Similarly, scholars of health misinformation could be a significant source of tactical insight, given the recent Measles outbreak and its relation to misinformation regarding vaccinations. Our approach is interdisciplinary, but our framework is information security.

6 Building a Usable Misinformation Framework

To create AMITT, we examined the data from two complementary directions: bottom-up from the incidents; and top-down from other frameworks that are used to plan activities similar to misinformation campaigns (such as *psyops* and *advertising*). The multiplicity of different types of incidents each have potentially very different stages, routes through them, feedback loops and dependencies.

We designed for coverage rather than perfection and focus on questions such as:

- What are the atomic actions in propaganda attacks?
- How do actions combine to form larger events, including more complex actions and attacks?
- How do the instances of attacks and actions combine to form campaigns?

6.1 Methodology

We analyze known misinformation incidents to identify their components and then place those components into a framework (*ATT&CK*) commonly used to describe information security incidents. The outputs from the working group include a misinfosec threat matrix designed for use by *blue teams* when considering options for defense and counter-attack, and by *red teams* when anticipating future attack types.

This process is iterative and collaborative. MisinfosecWG refines these strawmen into more detailed TTP descriptions and recommendations as options are tested or eliminated. Furthermore, we define major terms at focal points on the scale, with an emphasis on descriptive or procedural rigor.

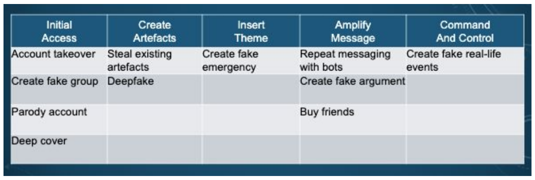
One of the operating assumptions of MisinfosecWG is that social and cognitive factors can "scale up and down"—facilitating some definitional and procedural crossover in both the construction of a framework for understanding these attacks and in their detection.

6.2 Keeping Compatibility with Existing Infosec Tools

6.2.1 The ATT&CK-based Strawman

We started by trying to adapt the ATT&CK framework for use in misinformation campaigns. The Mitre ATT&CK matrix is a foundational model, aligned with the Cyber Killchain, that provides a common language and tools. Concentrating on the ATT&CK model made sense when we started doing this work—it was detailed, well-supported, and had useful concepts such as technique-level responses and being able to group related techniques together under each stage.

We couldn’t get a direct fit between misinformation and the ATT&CK stages so a strawman set of stages was created for discussion (*Figure 6*), with the understanding that a more detailed bottom-up model would need to be developed later, and that the left-of-boom part of the model that wasn’t covered by ATT&CK was probably very valuable to responders.



Initial Access	Create Artefacts	Insert Theme	Amplify Message	Command And Control
Account takeover	Steal existing artefacts	Create fake emergency	Repeat messaging with bots	Create fake real-life events
Create fake group	Deepfake		Create fake argument	
Parody account			Buy friends	
Deep cover				

Figure 6: Initial ATT&CK-based strawman

6.3 Finding Techniques in Misinformation Incidents

6.3.1 Technique Analysis

We decomposed each incident into techniques, where a *technique* is defined as "how an adversary achieves a tactical objective by performing an action" [58]. To date, over 145 techniques have been identified across 63 incidents. Example techniques include:

- Establish metrics
- Create master narratives
- Create fake or imposter news sites
- Create fake experts (with impressive titles and factual sounding conclusions)
- Create fake websites
- Create funding campaigns
- Create hashtag
- Create fake video/ image
- Forge ('release') altered hacked documents
- Bait legitimate influencers (journalists, media, politicians)

Techniques are the activities that an attacker would do to execute an incident. Several techniques (e.g. amplification with botnets) are repeated across multiple incidents. Identifying these techniques and mapping them into the killchain are valuable to a responder because each could be potentially disrupted, and each could potentially leave traces online (e.g. planning activities might leave traces in search logs). Leveraging these insights increase the odds that defenders can intervene left of boom.

6.4 Mapping Techniques to Stages

Next, MisinfosecWG created a stage-based model by collecting together identical techniques (e.g. “manipulate online polls”); grouping binned techniques into stages; and ordering those stages temporally left-to-right (e.g. the stages that would have to be started earliest were on the far left). We identified a total of twelve stages, each with its own set of techniques.

Finally we cross-checked these “technique bins” with the model’s stages and techniques and reviewed the results in the context of each of the models listed above, looking for and filling gaps in the proposed model. One interesting result from this was that the New York Times model consisted of techniques, not stages. We used our tentative stage model to create a test incident (“convert a population of vaccine skeptics into antivax activists”), outlining counters for the test incident and some of the more common techniques, to test whether the model could potentially be used in a blue team context.

6.5 Procedures: Finding New Forms of Attack

Once we understand the techniques, we can use them as components in new combinations to describe threat procedures that we might not have considered before. A checklist of threats and best practices creates the space necessary to think more strategically about the misinformation environment and to balance institutional needs in the context of well-tested security principles.

6.6 Model Cleanup

The model evolves as it is discussed and used. First we grouped stages into phases (planning, preparation, execution, evaluation), and evaluated whether each phase was complete (e.g. wasn’t missing any stages). During this cleanup, we subsumed some stages into others (e.g. *Search Optimization* was subsumed into *Pump Priming*); and split some other stages apart (e.g. the *Planning* phase was split into *Strategic* and *Tactical* stages).

Next we separated the tasks in each stage from their techniques. A *task* is what you do and a *technique* is how you do it. This reduced the number of techniques in our model to 54. The resulting model, which we call AMITT, is described in the next section.

7 The AMITT Misinformation Framework

7.1 What AMITT is and Where to Find it

AMITT (*Adversarial Misinformation and Influence Tactics and Techniques*) is a framework designed for rapidly describing and understanding disinformation incidents. AMITT is part of misinfosec—work on adapting information security (infosec) practices to help track and counter misinformation—and is designed as far as possible to fit existing infosec practices and tools.

AMITT's style is based on the MITRE ATT&CK framework [36][48][98]; we're working on generating STIX templates [49][90] for all its objects so AMITT messages can be passed between ISAOs and similar bodies using infosec transport standards like TAXII.

AMITT is a living standard. The current model was built from an analysis of 22 incidents; the techniques listed in it are a subset of the ones used in misinformation incidents. The model will evolve as it's tested and used. We still have refinement work to do and expect testing to produce new insights. Additionally, we anticipate name changes and other recastings as techniques become familiar to researchers and responders.

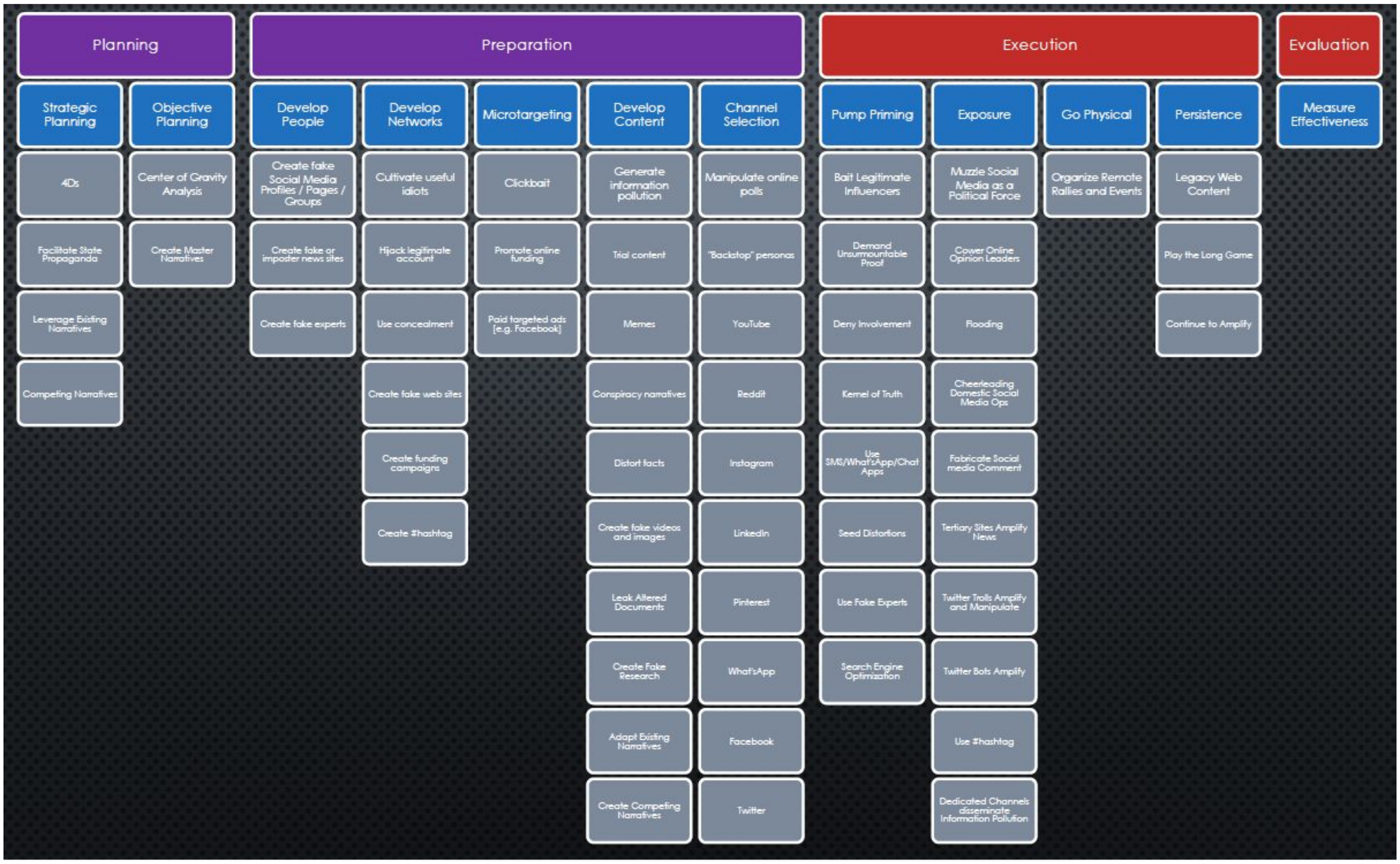
The latest version of AMITT is stored in our Github repository [89]

- https://github.com/misinfosecproject/amitt_framework

7.2 AMITT Components: Phases, Tactics, Techniques

Figure 7 shows a recent version of AMITT's framework diagram. The major components of the framework are:

- **Phases:** higher-level groupings of tactics, created so we could check we didn't miss anything. The tactics below each phase belong to that phase. (top row: purple and red boxes)
- **Tactics:** stages that someone running a misinformation incident is likely to use. (second row: blue boxes)
- **Techniques:** activities that an incident creator might use at each stage. The techniques below each tactic belong to that tactic. (all other rows: grey boxes):
- **Tasks:** things that need to be done at each stage. (not shown):



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
4Ds	Center of Gravity Analysis	Create fake Social Media Profiles / Pages / Groups	Cultivate useful idiots	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		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			

Figure 7: AMITT misinformation Framework

The language and style of AMITT is taken from the ATT&CK framework. The diagram (*Figure 7*) is read left-to-right in time, with the entities to the left typically (but not necessarily) happening earlier in an incident. The phases are separated into *left-of-boom* (purple) and *right-of-boom* (red). Left and right are used as a metaphor for the stage diagram, meaning *before* and *after,* respectively. The *boom* in a misinfosec context is either public exposure of a message payload on an otherwise measurable cognitive or social impact on the threat surface.

7.3 Phases

7.3.1 Planning

The planning phase is split into strategic and objective planning stages. The strategic planning stage contains ways (objectives), means (actions), ends (strategic plans) analysis. The techniques in each stage include strategic planning around the 4Ds (from [55]), leveraging existing narratives and creating competing narratives; objective planning includes center of gravity analysis (demographics etc) and creating master narratives to create artifacts around.

]

7.3.2 Preparation

The AMITT preparation phase includes stages to develop people, networks and content, microtargeting and channel selection. This includes a mixture of creating new “fake” accounts and sites, and leveraging existing ones for misinformation use. The preparation stage is still left of the “boom.”

7.3.3 Execution

The AMITT execution phase moves from a pump-priming stage (preparing to reach a larger audience by pushing a message to influencers, or conversely by dampening their ability to push a message) to exposing an audience to the narrative, with an optional move to physical actions, followed by activities that help the narrative persist.

7.3.4 Evaluation

The AMITT evaluation phase is used to measure the effectiveness of an incident, to help make future incidents more effective.

7.4 Describing an Incident with AMITT

7.5 Adding to AMITT

As AMITT is used—as new incidents arrive, as incident creators adapt to conditions including new response techniques and as more people see the model—new techniques, tasks (and possibly tactics) will be identified and existing ones will require modifications.

You can request changes to AMITT using our github repository [89]:

- Submit an issue; or
- Submit a pull request

Either way, our Github repo is the right place to add suggestions and thoughts.

7.6 Adapting AMITT

If you want to create your own versions of AMITT, or adapt it to a specific need, we’ve released the github repository under a CC-by-4.0 license; we’d also love it if you’d tell us how you adapted AMITT, in case there are things we’ve missed and need to add to the main models.

7.6.1 The AMITT Codebase

AMITT comes with code to generate a set of phase, tactic, technique, task and incident datasheets and associated matrices and lists from a simple excel spreadsheet. Once generated, these sheets contain areas that can be hand-edited as needed. As the project matures, these assets will become more web-interactive and sophisticated [89].

8 Misinformation Reporting and Response Protocols

8.1 Messaging Standards

Infosec messaging standards for sharing information about threats include STIX/TAXII and MISP.

8.1.1 STIX and TAXII formats

STIX and TAXII are described by their curators as follows.

TAXII™, the Trusted Automated eXchange of Indicator Information: TAXII defines a set of services and message exchanges that, when implemented, enable sharing of actionable cyber threat information across organizational, product line and service boundaries. TAXII is not an information sharing program itself and does not define trust agreements, governance, or other non-technical aspects of collaboration. Instead, TAXII empowers organizations to share the information they choose with the partners they choose [99][100].

STIX™, the Structured Threat Information eXpression: STIX is a collaborative effort to develop a standardized, structured language to represent cyber threat information. The STIX framework intends to convey the full range of potential cyber threat data elements and strives to be as expressive, flexible, extensible, automatable, and human-readable as possible. All interested parties are welcome to participate in evolving STIX as part of its collaborative community [83][84][90].

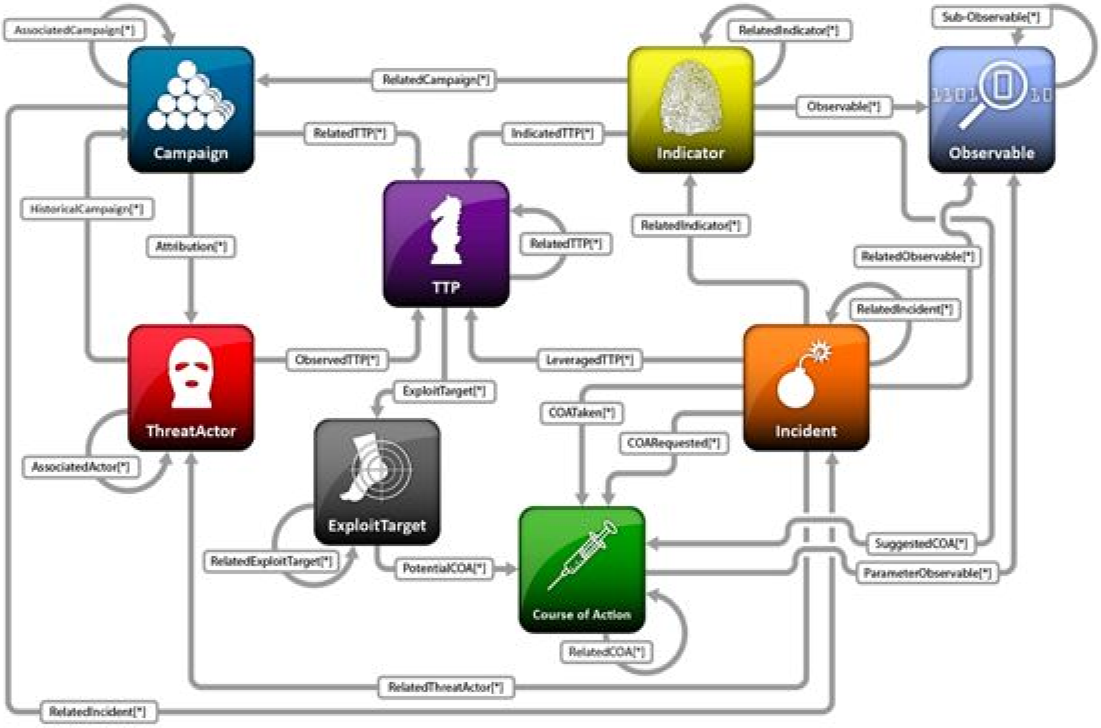


Figure 8: STIX model, used by government and industry [90]

8.1.2 Mapping from STIX Infosec to STIX Misinformation Standards

The strategy level is where we attack and respond. Each community has vulnerabilities which get exploited using TTPs, which leave artifacts online. When these are found, the responders create a report to share with other response groups. Responders who find these vulnerabilities can create reports that efficiently share their findings with other response groups. With shared standards, these reports are easy to generate and process and can serve as a roadmap for collective action.

Misinformation STIX	**Description**	**Category**	**Infosec STIX**
Report	Communication to other responders	Communication	Report
Campaign	An ongoing, persistent misiinformation attack	Strategy	Campaign
Incident	A short timeline misinformation attack focused on a particular time, place or event	Strategy	N/A (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	Artifacts 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

Panel 5: STIX Attributes for Misinformation Incidents

8.1.3 MISP Formats

MISP (Malware Information Sharing Platform) is an open source threat intelligence platform for sharing, storing and correlating *Indicators of Compromise* of targeted attacks, threat intelligence, financial fraud information, vulnerability information or even counter-terrorism information. It is intended to store, share, collaborate on cyber security indicators, malware analysis, but also to use the IoCs and information to detect and prevent attacks, frauds or threats against ICT infrastructures, organisations or people [85][86][87].

9 Use Cases

The AMITT framework will give misinformation responders the ability to transfer other information security principles to the misinformation sphere, identify gaps in known attack types, plan defenses and countermoves to common components, assess tools and mechanisms, build an information security style alert structure (cf US-CERT) and plan defenses for the types of large-scale adaptive threats that machine learning and other automation make possible [39][40][76].

Most offensive computer network operations are based upon misinformation. This similarity should aid in our task: network intruders want their targets to make decisions or take actions advantageous to them based on information that's shown, hidden, altered or destroyed. For example, *STUXNET* allegedly hid information about the true state of centrifuges from operators, enabling them to make an incorrect decision that no action needed to be taken.

In infosec, an organized taxonomy of attack and defense techniques allows operators to apply well-tested responses to familiar attack patterns, and learn from both successful and failed attacks. For misinformation, these interventions are likely to be drawn from various disciplines including sociology and psychology. For example, one possible intervention to a misinformation campaign is to push new information that draws opinion away from the goal of the original misinformation.

More campaign-aware approaches might seek to preemptively inoculate a vulnerable population against messaging known to be consistent with an attackers' objectives as discovered at earlier stages. At the campaign scale, interventions will often need to be prepared in advance for deployment in specific, measurable contexts. Examples include: the Macron teams' preparation in the 2017 French elections for the reuse of a 2016 US Presidential election technique (releasing and amplifying information from leaked political emails); the preparation of inoculating narratives; or the priming of strategic master narratives.

9.1 Red-Team, Blue-Team Exercises

To build a good defense, you need to understand your threat surface and the types of attacks that are likely to occur on it. The best way to understand attacks is to attack; in information security this is done through simulated attacks, where a *red team* attacks the systems of a defending *blue team*. These exercises typically expose previously-unseen system vulnerabilities.

Brundage et al [39] outlines a first red team playbook for misinformation, with common examples of: actors, targets, payloads, objectives, automation, and techniques. Extending this with the misinfosec work gives us a detailed catalogue of attack types, allowing platform defenders and autonomous blue team actors to test and stage effective countermeasures.

9.2 Alerting and Defense

In misinformation research, there's been a lot of "admiring the problem" and not so much action. But action does exist. You just have to look for it. Financially-motivated misinformation has been targeted at the *architecture* and *passive defense* levels.

But response against power-based misinformation generally needs to be faster, to be more coordinated, and to take advantage of knowledge gained by many defenders across many different fields. Action there includes: daily bomb alerts at the new IRA building; Macron's team building email honeypots in case US election tactics were used against them too; the Baltic Elves joining up with local media; and botnet removal, but not real-time or across platforms. But we need a way to share the knowledge of what was tried and what worked with the people who can act at the time that they need to act.

One of the reasons for building misinformation frameworks is that we can start to describe different levels of misinformation response in them:

- At the tactic level, we can develop a Courses of Action matrix. (It's still too early in AMITT's development for this)
- At the technique level, we can collate suggested and known counters to each technique

Absent a comprehensive counter-influence strategy from major government and intelligence players, the responsibility for misinformation management currently falls on individual persons and institutions. This mirrors the history of the information security field, which created bodies like the US-CERT organisation: the US government body that coordinates defenses and responses to cyber attacks.

A similar body for sharing and alerting may be needed here. US-CERT's work includes threat monitoring and analysis, information sharing, analytics, operations, communications and international partnerships. Its outputs include a current activity list, monthly active summaries, alerts, notes and tips and security publications.

These activities and outputs map well but not exactly to misinformation (different responses and connections are needed). And US-CERT already has a sister organization, ICS-CERT, which covers security of industrial control systems. We don't yet know which organizational roles will be the end user for a new body's product. Whoever is responsible for the adoption, deployment and enforcement of these practices, will probably require help from security professionals.

9.3 Counterattack (and its Limitations)

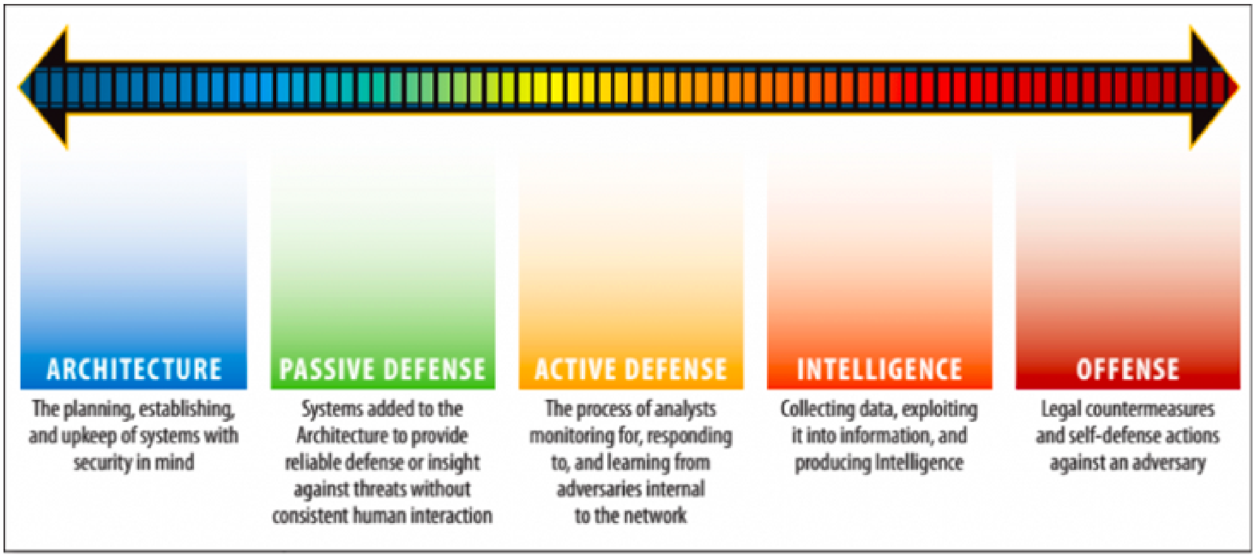

ARCHITECTURE
The planning, establishing, and upkeep of systems with security in mind
PASSIVE DEFENSE
Systems added to the Architecture to provide reliable defense or insight against threats without consistent human interaction
ACTIVE DEFENSE
The process of analysts monitoring for, responding to, and learning from adversaries internal to the network
INTELLIGENCE
Collecting data, exploiting it into information, and producing Intelligence
OFFENSE
Legal countermeasures and self-defense actions against an adversary

Figure 9: SANS sliding scale of cyber security

Democracies face structural disadvantages relative to the producers of misinformation. Clint Watts [41] cites the Kremlin's strategic edge from Russia's cybercrime underworld and the plausible deniability it gives; the US and its allies don't have this advantage.

> *Ultimately, America's problem in counterinfluence is that we don't know what to say … During the Cold War, the United States promoted democracy and democratic values. But today the United States doesn't appear to know what it wants. Quite simply, if America doesn't have its feet on the ground, then it can't push back at those challenging us.* [41]

This seems right, but generational. In the meantime, we need a plan. One that doesn't sacrifice what we believe in. This is hard. We believe that transparency is key for democratic actors. Knowing who is actually delivering the message goes a long way. Of course, there are cases in which we want to protect the source (e.g. dissidents in an autocracy), but those are the exception rather than the rule. The framework we propose is agnostic to locality and describes components; it is up to the "local populace" to decide acceptable countermeasures. In other words, part of the remedy is democratic participation itself.

10 Conclusions and next steps

We've started the work of adapting information security frameworks for misinformation tracking and counters, but there is much work still to do. The information security field has decades of experience that we can draw on in our work, but there have been enough differences between the fields for us to create a new framework, albeit one based on ATT&CK. Challenges we anticipate in this work include: *epistemology*, working across multiple very different fields of research, defining and naming different levels and stages of "attack"; *advocacy*, persuading people that information security frameworks are already about human influence systems; and *legitimacy*, legal and ethical constraints on response.

10.1 Challenges

While information security attacks are firmly rooted in the quantitative field of computer science, influence campaigns are, by necessity, rooted in the qualitative fields of sociology and psychology. The linking of quantitative and qualitative fields of science has always been epistemically precarious. Additionally, any attempt to develop an overarching and generalized framework will necessarily omit details. No overarching framework will ever be completely accurate in all situations.

The AMITT framework provides an ontology for influence campaigns whether or not they are executed exclusively in the cyber domain. As with any complex system, there will be an emergence of properties which is greater than the sum of its parts. AMITT describes influence from the view point of tactics, techniques, and procedures (TTPs) without consideration for the intent, morality, or legality of such actions. Analysis of morality, legality, and intent are beyond the scope of this work.

10.2 Next Steps for the MisinfosecWG

10.2.1 Testing/ refining the framework

We need to test the framework against new incidents—both historical incidents that we haven't included in it, and new incidents as they emerge. Part of this work is to find existing response populations who could use the framework, and determine the training and adaptations they need to be able to use it themselves. This will make the framework more useful both to them, and to future potential users.

10.2.2 Blue Team Playbook

We have phases, stages and techniques listed. The next part of the work is to detail the techniques, and the potential responses to them, including defensive measures against them at each stage. AMITT will form the basis of a blue team playbook for misinformation incident response. The use of the playbook will inform testing and refining of the framework.

10.2.3 Support to emerging Response Centers

10.3 Thanks

MisinfosecWG would like to thank:

- The Credibility Coalition for not blinking when we said “hey, we’ve got this idea that merges hackers and misinformation and it’s going to be fine”
- The Newmark Foundation for the funds that meant we could hold the CredConX Atlanta workshop that turned an idea into a working framework, and engage a lead contributor to accelerate its development
- Misinfosec, our sister group, for advice, links, and assistance whenever we needed it

11 References

11.1 MisinfosecWG Outputs to Date

11.1.1 Research Papers and Blog Posts

- Walker, Terp, Breuer, Crooks, Misinfosec: Applying Information Security Paradigms to Misinformation Campaigns, Companion proceedings of the 2019 World Wide Web Conference, San Francisco USA, May 13-17 2019
- Terp, Misinformation has Stages: Now we just need to work out what the stages should be…, Medium, May 13 2019
- Gray, Terp, The MisinfoSec Framework Takes Shape: Misinformation, Stages, Techniques and Responses. Developing the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) framework, Medium, Jun 19 2019
- Gray, Terp, BRIMs2019_Misinformation_We're Four Steps Behind Its Creators

11.1.2 Presentations

- Sofwerx, (December 2018)
- Cyber Defense and Network Security (CDANS) London, (December 2018)
- SOCOM, (February 2019)
- NYU, (February 2019)
- CUNY, (February 2019)
- CanSecWest, (March 2019)
- People Centered Internet, (March 2019)
- Georgetown University, (March 2019)
- Oktane, (April 2019)
- Asilomar Microcomputer Workshop, (April 2019)
- WWW, (May 2019)
- AI World Government, (May 2019)
- WebIT, Sofia, (May 2019)
- CogX19, (June 2019)
- NATO (June 2019)
- Bournemouth University (June 2019)
- UK Information Commissioners Office (June 2019)
- BRIMs (July 2019)
- BlackHat, (August 2019)
- Security BSides Las Vegas, (August 2019)
- Narwhal, (August 2019)
- 3rd Annual International Information Sharing Conference (August 2019)
- Harvard Disinformation Workshop (October 2019)
- Trust & Truth Conference (October 2019)

- Mitre ATT&CKcon 2.0 (October 2019)

11.1.2.1 Pending

- HIICS (January 2020)

11.2 Referenced Documents

[0] Credibility Coalition. Collected 2019. https://credibilitycoalition.org/

[1] Y. Benkler, R. Farris and H. Roberts, Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics, Oxford: Oxford University Press, 2018.

[2] Smith and Banic, "Fake News: How a partying Macedonian teen earns thousands publishing lies," NBC News, 2016.

[3] R. DiResta, "The Information War is on. Are we Ready for it?” Wired, 2018.

[4] A. Petrov, "Modeling Position Selection by Individuals during Information Warfare in Society," *Mathematical Models and Computer Simulations,* pp. 401-8, 2016.

[5] S. Oates, "When Media Worlds Collide: Using Media Model Theory to Understand How Russia Spreads Disinformation in the United States," in *American Political Science Association 2018 Annual Meeting*, Boston, MA, 2018.

[6] S. Oates, J. B. Barrow and B. Foster, "From Network to Narrative: Understanding the Nature and Trajectory of Russian Disinformation in the U.S. News.," in *International Journal of Press/Politics Conference*, Oxford, UK, 2018.

[7] A. Field, D. Kliger, S. Wintner, J. Pan, D. Jurafsky and Y. Tsvetkov, "Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies.," EMNLP, 2018.

[8] T. Boucher, "Adversarial social media tactics," Medium, 2018.

[9] S. J. Terp, "Security frameworks for misinformation," 2018.

[10] S. J. Terp, "Practical influence operations," 2018.

[11] S. J. Terp, "Social engineering at scale," 2018.

[12] Stubbs, "Exclusive Iran based political influence operation - bigger, persistent, global," Reuters, 2018.

[13] F. Mortola, "Disinformation through fabricated news site," 2018.

[14] Foster, "Influence Operations Targeting the 2018 U.S. Midterms: What are we seeing? What are we not?," 2018.

[15] Ehmke, "Influencer Vaccine: Identifying Information Operations Infrastructure," 2018.

[16] Kuhr, "Leveraging threat intel disinformation campaigns to defeat attribution," 27 February 2017. [Online]. https://www.synack.com/2017/02/27/shmoocon-2017-recap-election-hackers-vs-threat-intel-attribution

[17] Landau, "Cybersecurity: time for a new definition," 2018.

[18] Rogers, "Fake news as an information security problem," May 2018.

[19] Rogers, Director, *Fake news as an information security problem.* [Film]. 2018.

[20] D. Gordon, "Defending the Indefensible: A New Strategy for Stopping Information Operations." 2018.

[21] Grugq., "Lessons in Cyber: Influence Operations," 2018.

[22] A. Zegart, H. Lin, T. Fingar, N. Persily and L. Ross, "Cyber-Enabled Information and Influence Warfare and Manipulation: Understanding Problems, Developing Solutions," 2017.

[23] H. Lin and J. Kerr, "On Cyber-Enabled Information/Influence Warfare and Manipulation," 2017.

[24] H. Lin, "Developing Responses to Cyber-Enabled Information Warfare and Influence Operations," 2018.

[25] FreedomHouse., "Manipulating Social Media to Undermine Democracy," 2018.

[26] S. Zannettou, T. Caulfield, W. Setzer, M. Sirivianos, G. Stringhini and J. Blackburn, "Who Let The Trolls Out? Towards Understanding State-Sponsored Trolls," 2018.

[27] D. D. Kirkpatrick, "Signs of Russian Meddling in Brexit Referendum," 2017.

[28] D. O'Sullivan and D. Byers, "Exclusive: Fake black activist accounts linked to Russian government," 2017.

[29] V. Joler, M. Jovanović and A. Petrovski, "Mapping and quantifying political information warfare Part 1 : Propaganda, domination & attacks on online media," 2016.

[30] M. Rosenberg, N. Confessore and C. Cadwalladr, "How Trump Consultants Exploited the Facebook Data of Millions," 2018.

[31] A. Nossiter, D. E. Sanger and N. Perlroth, "Hackers Came, but the French Were Prepared," 2017.

[32] H. G. Frankfurt, On Bullshit, Princeton, NJ: Princeton Univerity Press, 2005.

[33] DoD, "Joint Publication 3-13: Information Operations, February 13, 2006," 2006.

[34] Cyberpedia, "How to Break the Cyber Attack Lifecycle." Collected 2019.

[35] B. Ninmo, "The 4 Ds of Propaganda: Dismiss, Distort, Distract, Dismay," 2017.

[36] MITRE, "Att&ck matrix for enterprise," 2019. [Online]. https://attack.mitre.org/

[37] "Post-Exploit Threat Modeling with ATT&CK," 2016.

[38] K. H. Jamieson, Cyberwar: How Russian Hackers and Trolls Helped Elect a President - What We Don't, Can't, and Do Know, Oxford. UK: Oxford University Press, 2018.

[39] M. Brundage, A. Shahar, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, H. Anderson, H. Roff, G. C. Allen, J. Steinhardt and C. Flynn, "The Malicious Use of Artificial Intelligence:," Archiv, 2018.

[40] M. Brockman, "Data-Driven Propaganda as a Subset of Adversarial Examples," 2018.

[41] C. Watts, Messing with the Enemy: Surviving in a Social Media World of Hackers, Terrorists, Russians, and Fake News, New York, NY: Harper Collins, 2018.

[42] C.R. Walker, S.J. Terp, P.C. Breuer, C.L. Crooks, “Misinfosec: applying information security paradigms to misinformation campaigns”, Companion proceedings of the 2019 World Wide Web Conference, San Francisco USA, May 13-17 2019

[43] Insikt Group. “Beyond Hybrid War: How China Exploits Social Media to Sway American Opinion” Recorded Future. March, 2019.

[44] EPRS. “Polarisation and the use of technology in political campaigns and communication” European Parliamentary Research Service. March, 2019.

[45] Smith and Banic, "Fake News: How a partying Macedonian teen earns thousands publishing lies," NBC News, 2016.

[46] D. Bernardi, P. Cheong, C. Lundry, and S. Ruston, “Narrative Landmines: Rumors, Islamist Extremism, and the Struggle for Strategic Influence.” Rutgers University Press 2012

[47] P Pols, “The Unified Kill Chain: Designing a Unified Kill Chain for analyzing, comparing and defending against cyber attacks”, Cyber Security Academy, 2017

[48] MITRE ATT&CK https://attack.mitre.org/ Collected 2019.

[49] STIX™ Version 2.0. Part 1: STIX Core Concepts. Edited by Rich Piazza, John Wunder, and Bret Jordan. 19 July 2017. OASIS Committee Specification 01.

[50] “What is AB testing?” https://vwo.com/ab-testing/ Collected 2019.

[51] US Department of Justice, “Report of the Attorney General’s Cyber Digital Task Force”, July 2018

[52] C. Watts, “Five Generations of Online Manipulation: The Evolution of Advanced Persistent Manipulators”, Foreign Policy Research Institute

[53] Personal communication

[54] Personal communication

[55] B. Nimmo, “Anatomy of an Info-War: How Russia’s Propaganda Machine Works, and How to Counter It”, CEPolicy.org, 15 May 2015

[56] Hutchins, “Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains”, 6th international conference on iwarfare and security, 2011

[57] L. M. Neudert, "Computational Propaganda in Germany: A Cautionary Tale" University of Oxford 2017

[58] Center for International Relations, “Information Warfare on the Internet.” 2017

[59] Atlantic Council, “Defining Russian Election Interference: An Analysis of Select 2014 to 2018 Cyber Enabled Incidents.” 2018

[60] FireEye Intelligence, “Suspected Iranian Influence Operation Leverages Network of Inauthentic News Sites & Social Media.” 2018

[61] Presidential Policy Directive — Critical Infrastructure Security and Resilience, PPD-21

[62] M.S Chan, C.R Jones, K.H Jamieson, & D Albarracin. “Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation” Association for Psychological Science. Vol 28(11). 2017.

[63] “Cyberspace operations”, Joint Publication 3-12

[64] Acquia. “Getting Started: Collaborative development with Git” Acquia Developers. 2013

[65] ISAO Enrollment Page. Collected 2019. https://www.isao.org/information-sharing-groups/enrollment/

[66] DHS Interagency Security Committee. “Presidential Policy Directive 21 Implementation: An Interagency Security Committee White Paper” 2015

[67] ISAOs at DHS. Collected 2019. https://www.dhs.gov/cisa/information-sharing-and-analysis-organizations-isaos

[68] B. Schneier “Information Attacks against Democracies” Schneier on Security. 2018

[69] N. Schmidle. “The Digital Vigilantes who Hach Back.” The New Yorker. April 30, 2018

[70] C. Watts. “Advanced Persistent Manipulators, Part Two: Intelligence-led Social Media Defense.” Alliance for Securing Democracy. April 24, 2019

[71] J.C Ong and V.A Cabanes. “Architects of Networked Disinformation” The Newton Tech4Dev Network. February, 2018.

[72] H. Farrell and B. Schneier “Defending Democratic Mechanisms and Institutions against Information Attacks” Shneier on Security, 2019

[73] R. DiResta. “We’ve diagnosed the disinformation problem. Now, what’s the prescription?” Defusing Disinfo, January 23, 2019.

[74] “The instruments of national power” The Lightning Press. Collected 2019.

[75] H. Farrell & B. Schneier "Common-Knowledge Attacks on Democracy" Berkman Klein Center for Internet and Society. Harvard University. October, 2018

[76] S.C. Wooley & P.N Howard (eds) Computational Propagnda. Oxford. 2019

[77] C. Paul & M. Matthews. "The Russian Firehose of Falsehood Propaganda Model" Rand. 2016

[78] Y Sodeoka. "Why Technology Favors Tyranny" The Atlantic. October, 2018.

[79] K. M. Carley, G. Cervone, N. Agarwal, H Liu. "Social Cyber Security" 2018

[80] J Stray. "Institutional Counter-disinformation Strategies in a Networked Democracy" 2019.

[81] Interpol. "Investigative support for cybercrime" Collected 2019. https://www.interpol.int/en/Crimes/Cybercrime/Investigative-support-for-cybercrime

[82] U. Saiidi. "Inside Interpol's Singapore cybercrime-fighting complex." CNBC. May 17, 2017.

[83] MITRE. "Introduction to STIX." Collected 2019. https://oasis-open.github.io/cti-documentation/stix/intro

[84] MITRE. STIX™ Version 2.0. Part 2: STIX Objects. Working Draft 03. 19 June 2017.

[85] MISP - Open Source Threat Intelligence Platform & Open Standards For Threat Information Sharing. Collected 2019.
https://www.misp-project.org/index.html

[86] MISP Categories and Types. Collected 2019.
https://www.circl.lu/doc/misp/categories-and-types/

[87] MISP Feed OSINT. Collected 2019.
http://www.botvrij.eu/data/feed-osint/

[88] DHS Interagency Security Committee. "Presidential Policy Directive 21 Implementation: An Interagency Security Committee White Paper" February, 2015.

[89] AMITT Project on Github. Collected 2019.
https://github.com/misinfosecproject/amitt_framework

[90] US-CERT. "Information Sharing Specifications for Cybersecurity." Collected 2019.
https://www.us-cert.gov/Information-Sharing-Specifications-Cybersecurity

[91] MisinfosecWG Mission Statement, December 2018.
https://github.com/credcoalition/community-site/wiki/MisinfoSec-(The-Intersection-of-Misinformation-and-InfoSec)

[92] J. Gray and S.J. Terp. "The MisinfoSec Framework Takes Shape: Misinformation, Stages, Techniques and Responses: Developing the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) framework" Medium. June 19, 2019.

[93] Rand Corporation. Topics: Information operations. Collected 2019.

[94] Technopedia: Information Security (IS). Collected 2019.

[95] L. Steckman and A.R. Mallory. "Combining Narratology and Psychology to Examine Multinational Cultural Motivators, Expression, and Perceptions A National Academies of Science Research Solutions Whitepaper." MITRE Corporation. Case Number 17-2382. June 2017.

[96] Marvelous AI: Story Arc Alpha. 2019. https://insights.marvelous.ai

[97] P. M. A. Linebarger. *Psychological Warfare*. LoC CC No: 48-1799. 1948.

[98] MITRE ATT&CK Project on Github. Collected 2019. https://github.com/mitre-attack/attack-website/

[99] Trusted Automated eXchange of Indicator Information (TAXII™) 1.x Archive Website. Collected 2019. http://taxiiproject.github.io/

[100] STIX and TAXII standards. Cyber Threat Intelligence Technical Committee. Collected 2019. https://oasis-open.github.io/cti-documentation/