Misinformation: We're Four Steps Behind Its Creators

John F Gray[1] and Sara-Jayne Terp[2]

[1] Credibility Coalition Misinfosec Working Group

john@mentionmapp.com

[2] Credibility Coalition Misinfosec Working Group

sarajterp@gmail.com

Abstract. The Credibility Coalition MisinfoSec Working Group's earlier work described misinformation incidents as a series of tactical stages. This paper discusses the work needed to determine those stages, including whether we need more than one model for misinformation. We describe our methodology and work on which stages are appropriate for misinformation tracking. The structure and propagation patterns of misinformation incidents have many similarities to those seen in information security. The MisinfoSec Working Group has analyzed the similarities and adapted information security standards (e.g. ATT&CK) to deliver Version 1.0 of the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) framework on July 27, 2019. This framework presents better ways to describe, identify, disrupt and counter the techniques, tactics, and procedures (TTPs) used in misinformation incidents and campaigns. The AMITT framework will help better analyze, communicate and counter activities "left-of-boom" (an explosives term meaning the time before an attack – a period when you still have time to prepare and avert a crisis). This new information-sharing framework will assist misinformation responders and response organizations as an effective component in reducing the further erosion of truth and trust in our democratic institutions. On August 27, 2019, we published a six-month project report - "Building standards for misinfosec. Applying information security principles to misinformation response."

Keywords: Misinformation, Information Security, Framework

1. Introduction

We use “misinformation incident” to refer to the deliberate promotion of false, misleading or misattributed information. 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. The structure and propagation patterns of misinformation incidents have many similarities to those seen in information security. The Credibility Coalition’s Misinfosec Working Group (“MisinfosecWG”) is analyzing those similarities, including adapting information security framework standards to give better ways to describe, identify, disrupt and counter the techniques, tactics, and procedures (TTPs) used in misinformation incidents.

The process [1], began with building a strawman framework (figure 1), based on the Mitre ATT&CK framework (used by the infosec community to share information about incidents), and described how we were populating it by analyzing known misinformation incidents.

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 1: ATT&CK-based strawman

Concentrating on the ATT&CK framework made sense when we started doing this work—it was detailed, well-supported, had useful concepts such as grouping related techniques together under each stage, and covered the artifacts (messages, botnets etc.) seen by a system defender. But even with data, we were still discussing what the stages of the misinformation model should be (and whether there was one model or a family of models), and ATT&CK doesn’t cover the ‘left of boom’ (refers to the moments before an explosion or attack – a period when you still have time to prepare and avert a crisis) work that a misinformation incident creator does before releasing messages, images, etc. This leads us to map other potential frameworks to misinformation incidents. The following highlights the evolution of this process to create a new and unique framework for responding to misinformation campaigns.

2. Creating a master list of misinformation stages

We exclusively researched existing stage-based models (models that divide an incident into a sequence of stages, e.g. 'recon'). We found many models to choose from, but none of them were 'right enough' for general misinformation incidents. Figure 2 maps the stages in the models of most interest to us.

Marketing 1	Marketing 2	Cyber Killchain	Psyops phases	Justice Department	Bruce Schneier
		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 2: Comparing stage-based models

Central to this is the Cyber Killchain model (figure 3), which is the parent model of the ATT&CK framework. ATT&CK adds more detail to the last 3 stages of the Cyber Killchain: these "right of boom" stages happen after bad actors gain control of a network and start damaging it; the other cyber kill chain stages are "left-of-boom".

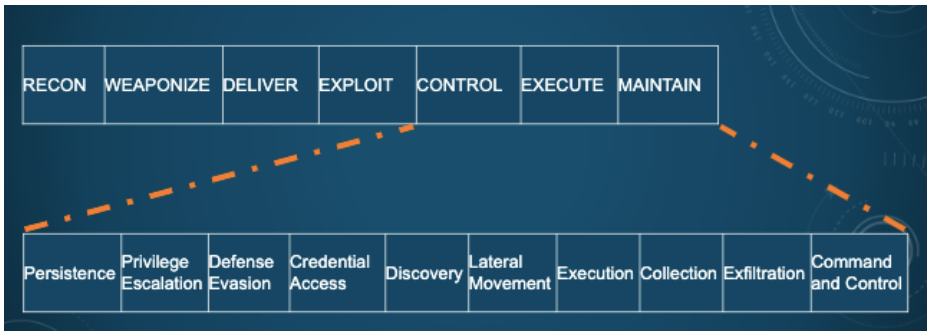

Figure 3: *Cyber Killchain stages (top), ATT&CK framework stages (bottom)*

Digital marketing funnels (figure 4) describe the 'customer journey' (the end consumer) of a marketing campaign, as moving from seeing an online brand message into taking an interest, then building a relationship with a brand/idea/ideology and finally advocating it to others. The point of view of this model leads us to consider the relationship between the people targeted by a misinformation incident, the people delivering it, and the people defending against it. Viewing this from the creator/attacker point of view for misinformation models, we can see how each attacker stage, including the ones less visible to a defender, can potentially be disrupted. Digital marketing could be useful in describing radicalization and including an advocacy stage: this mirrors other models' use of amplification and 'useful idiot' stage. Furthermore, the lends itself to adding the idea that an 'infected' node in the system isn't just repeating a message but might be or could become a command node too. Marketing funnels are "right of boom", so we've added marketing planning and production stages (market research, campaign design, content production) to see if they could be useful to describing and disrupting an attacker's game plan.

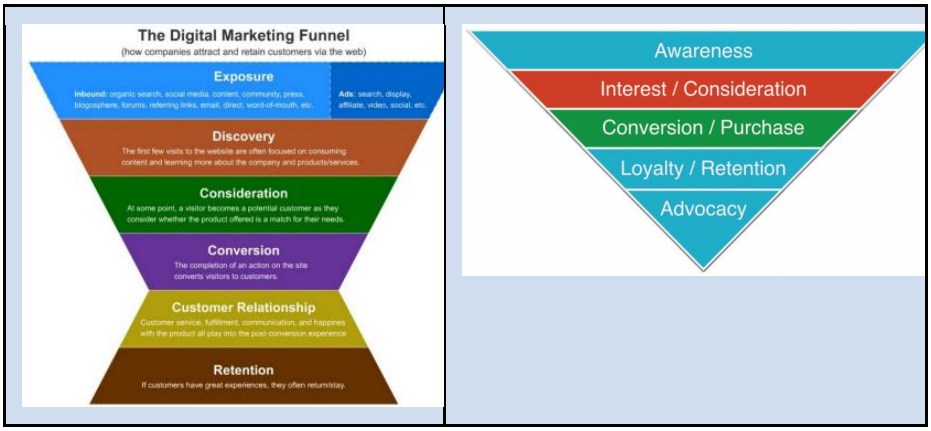

Figure 4: Digital marketing funnels [4] [5]

The Psyops model (figure 5) offers the point of view as an incident creator (or campaign creator - building a group of related incidents), with them controlling every stage, from planning through to evaluation, including human-hierarchy-aware procedures such as sign-offs and permissions, but with little visibility of end-consumer-specific considerations (these are bundled under "production, distribution, dissemination"). The evaluation stage is useful: one of the strengths of working at scale online is the ability to test hypotheses (such as content effectiveness and audience engagement) and adapt quickly at all stages. Additionally, when running a campaign, after-action reviews can be invaluable in learning and adjusting higher-level tactics (e.g. the list of stages, the target platforms, the most effective narrative styles and assets) between incidents.

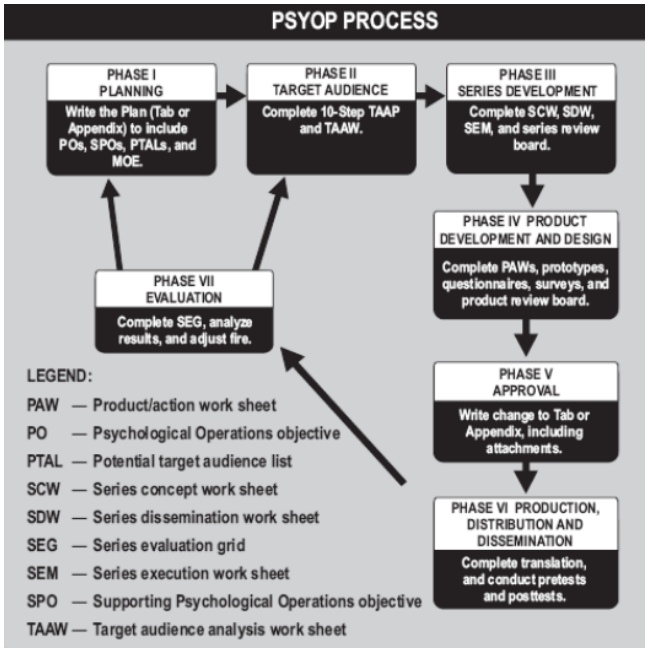

Figure 5 Psyops model [6]

The Department of Justice misinformation model (figure 6) clearly presents what each stage looks like from both the attacker and the defender points of view (the end consumer isn't of much interest here). It offers a helpful description of early IRA incidents yet is no longer current for recent incident types and doesn't include other state actors (such as China, Iran, Turkey, North Korea, Venezuela, and Saudi Arabia). This serves as a good example of

how we can create models that work well for some but not all the misinformation incidents that we've seen or expect to see.

Figure 2: The Malign Foreign Influence Campaign Cycle

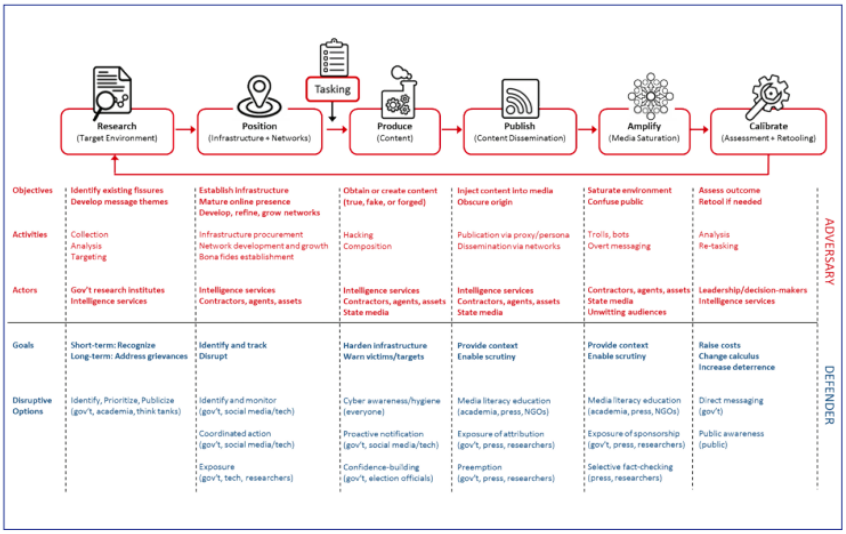

Figure 6: Department of Justice misinformation model [7]

There are other models. Ben Decker's model (figure 7) looks at misinformation campaigns as a series of handoffs between groups on different platforms, from the originators of the content to command and control signals, posting content to social media platforms, amplifying narratives with social media messages, to 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 offers a number of important considerations.

Figure 7: Ben Decker's misinformation model

3. Looking "left of boom"

Our research on current stage-based models helped us recognize that 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, the Credibility Coalition MisinfosecWG is focused on giving 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.

While our work is focused on creating a framework for responding to misinformation campaigns, we have identified that the "left of boom" stages could be the most valuable places to disrupt any incident. Using cyber kill-chain stages, we can recognize how adversaries intent on changing the beliefs of both a large audience or the strategically targeted few exploit the vulnerabilities found in these four operational phases:

- **Reconnaissance -** The attacker has the advantage here, with easy access to social space and OSINT data, combined with anonymity and deception making mass target information gathering and profiling cheap easy and low risk.
- **Weaponization -** There's a proliferation of free/inexpensive tools to create content (rumors, lies, outrages, conspiracies) and generate memes/images/audio and video (although it isn't vital to use originals). Historical Psyops principles still apply today, e.g. wrapping rumor &

innuendo in a grain of truth, using outrage, doubt, conspiracy and humor, and exploiting existing themes/seams and social polarities.

- **Delivery -** Can distribute to multiple platforms distribution as 1:1, 1:few or 1:many; platforms range from WhatsApp, Twitter, Tinder [8] to Facebook, YouTube, and Blackhat with search engine optimization (RT.com are masters at getting news at/near the top of news search rankings).
- **Exploitation -** bots amplify content to make it look popular/viral in metrics; trolls and "useful idiots" lay bait for journalists, politicians, business leaders, and the public. At the volume of supply, speed of consumption, and shallowness of engagement for much of the audience, sources are irrelevant, and verification is unwarranted particularly when it's feeding deeply entrenched human biases.

The Credibility Coalition working group efforts recognize that 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.

A framework needs to provide an ontology for influence campaigns whether 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. At this point, our work describes influence from the viewpoint of tactics, techniques, and procedures (TTPs) without consideration for the intent, morality, or legality of such actions. Analysis of morality, legality, and intent is beyond the scope of this work, as such creating left-of-boom responses that include disrupting, co-opting, denying and displacing are currently speculative. Applying these counters to the recon stage, for instance, we could deploy potential actions that include:

- **Disrupt** - build honeypots to help find the actors and trace them home; get an adversary to reveal themselves
- **Co-opt** - create and deploy personas with new narratives and information
- **Deny** - remove the narrative power in the space. Note that this doesn't equal censorship, but the removal of artificial positions of strength
- **Displace** - create models of community, identity, and trust that move bot- and troll-like behaviors away. Note that this doesn't mean purging anonymity.

4. Conclusions and Future Work

We have 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.

On July 27, 2019, The Credibility Coalition Misinfosec working group introduced Version 1.0 of the AMITT (Adversarial Misinformation and Influence Tactics and Techniques) framework and acknowledge that any attempt to develop an overarching and generalized framework will necessarily omit details. No overarching framework will ever be completely accurate in all situations.

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.

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 the testing and refining of the framework.

On August 27, 2019, we published our six-month project report - "Building standards for misinfosec. Applying information security principles to misinformation response."

The Misinfosec WG will continue working towards the joined-up responses with the goal to have these standards adopted by today's Information Sharing Analysis Organizations (ISAO's), and response centers.

5. References

1. Walker, C, .: Misinfosec: applying information security paradigms to misinformation campaigns. W3 Workshop on Misinformation (2019).
2. Schneier, B.: (2019).
3. Sterling, B.: The Hacker Crackdown (2019).
4. https://moz.com/blog/building-your-marketing-funnel-with-google-analytics
5. https://www.singlegrain.com/marketing-funnels/how-to-build-a-social-media-marketing-conversion-funnel/
6. https://2009-2017.state.gov/documents/organization/148419.pdf
7. https://www.justice.gov/ag/page/file/1076696/download
8. https://www.wired.co.uk/article/tinder-political-bots-jeremy-corbyn-labour

6. Biographical notes

John Gray is a co-chair of the Credibility Coalition Misinfosec Working Group and is the co-founder of Mentionmapp Analytics Inc. He's been actively researching and reporting on issues related to global disinformation since 2016. He's co-authored "#Kremlin: Using Hashtags to Analyze Russian Disinformation and Audience Engagement", and has collaborated with journalists and publications including the Reuters feature "Hatebook" (part of the series "Myanmar Burning" which was awarded the 2019 Pulitzer Prize for International Reporting). John has a Bachelor of Applied Science (Communications) and a B.A. (English) from Simon Fraser University.

"SJ" Terp is a data scientist, strategist, old-school AI researcher and community builder who focuses on complex business and social problems. She's currently working on the Global Disinformation Index (an independent disinformation rating system), and running a Credibility Coalition working group on the application of information security principles to misinformation; her previous work covers belief systems and situation awareness across many disciplines (including autonomous systems, intelligence analysis, crisis data, journalism, online advertising and political data science).