サイバー空間での影響力工作に対するアトリビューションの概要

齋藤孝道
（明治大学サイバーセキュリティ研究所所長・理工学部教授）
2023 年 10 月 5 日

要旨

SNS にバトルフィールドを広げる影響力工作の背後には誰がいるのか、その意図は何か。本稿は、サイバー空間での影響力工作におけるアトリビューションを概説し、これらを明らかにする方途について論じる。アトリビューションに関する概念や分析のサイクル、モデルを提示し、2023 年 G7 外相会議直前に展開されたキャンペーンを対象にモデルの実践を行う。また、意図の見積もり、データの入手、真の発信者の特定といった点から、サイバー空間に閉じた情報源を用いたアトリビューションの限界についても論じる。

はじめに

本稿では、IT やデータ分析などの素養のある読者を想定し、文献[1-4]及び筆者らが過去行った分析をベースに、サイバー空間での影響力工作におけるアトリビューション(以降、アトリビューション)を概説することを試みる。なお、本稿で使う用語は、それらの文献に基づく。また、特定を避けるため、一部曖昧な表現を使用していることは容赦願いたい。

影響力工作及びアトリビューション

影響力工作を、ここでは「国家間での競争(戦い)における情報戦の一種で、競争相手国の意思決定に影響を与え、ターゲットの行動の変容を促す一連の行為」とする。行動抑制などの行動変容に繋げることが特徴である。影響力工作には以下の目標がある。詳細は文献[1]を参照されたい。

① 社会的不和の誘発
② 政権政党の支持向上
③ 選挙介入・正統性低下
④ リクルート

特に、「社会的不和の誘発」は、敵対国家における世論の分断化などにより、その国の指導者の求心力を弱め、重大なことへの政治的決断を困難にすることが目的である。影響力工作は、20 世紀初頭に登場したラジオやその後のテレビの普及により高度化・体系化され、インターネットの普及により、そのバトルフィールドを SNS に広げた。本稿では、アクターが秘匿され実施されるケースにおけるアトリビューションを議論の対象とする。

次に、文献[3]に基づき、アトリビューションを定義する。ここでは「サイバー活動の背後に誰がいたのか、なぜそのようなことをしたのかを答えようとする分析プロセス」とする。一般に、国家安全保障の観点において、競争(敵対)関係にある国により与えられる脅威は競争国の能力と意図の積として導出される。よって、アトリビューションは、脅威分析の観点で、競争国における意図を見積もること(以降、意図見積)を一部包含する。

インテリジェンス・サイクル

アトリビューションを特定し要求者に提供するまでのプロセスを、本稿では、インテリジェンス・サイクルと見立て、以下のステップで定める。紙面の都合で一般的な説明は文献[2]などに譲る。本稿でのポイントに絞って補足する。

① **計画・指示（Planning & Direction）**：ターゲットを特定し、情報収集方針を定める。
② **収集（Collection）**：特定の情報源から必要な情報、特に、侵害指標（Indicator of Compromise: IoC）を収集する[1]。複数メディアで横断的に集めることが期待される。
③ **処理（Processing）**：処理は、選別、分類、評価および保管からなる。プロファイリング、クロノロジーの作成も含む[2]。
④ **分析・作成（Analysis & Production）**：情報を分析し、求められるインテリジェンスを作成する。
⑤ **配布（Dissemination）**：分析されたインテリジェンスを、関連するステークホルダーに配布する。
⑥ **評価（Evaluation）**：インテリジェンスの使用結果や品質を評価し、必要に応じてサイクルを回す。

「計画・指示」では、アトリビューションの分析ターゲットとするキャンペーンを定める[3]。その上で、どのようなデータをどの程度集めるのか、どのような手法で分析するのか、即時性を求めるかなどの方向性を定める。利用するツール、及び組織内での役割分担や投入コストも決める。一般に、IT インフラへのサイバー攻撃におけるアトリビューションの場合、攻撃があった事実をもってアトリビューションを行う。その一方、影響力工作におけるアトリビューションの場合、それが工作なのか、自然発生のバイラルなのかを切り分ける必要がある。アトリビューションは、ナラティブやミームにより当該キャンペーンを特定する。

「収集」では、当該キャンペーンに関して、SNS、ブログサイトやニュースサイトから、分析フェーズで必要なデータを集める。X（旧 Twitter）に関して採取項目は、当該キャンペーンのツイートの量、頻度、投稿時間（タイムゾーン、休暇期間などを考慮）、アカウントの作成時期や利用言語などが想定される。コピー＆ペーストのみでの投稿やハッシュタグのみでの投稿も採取する。また、機械学習を用いてボット判定や位置情報などを推定する場合、それらに必要なデータも採取する。併せて、クロノロジー作成のための情報も採取する。可能である場合、デジタル以外での情報として、関係者からの内部情報も入手する。

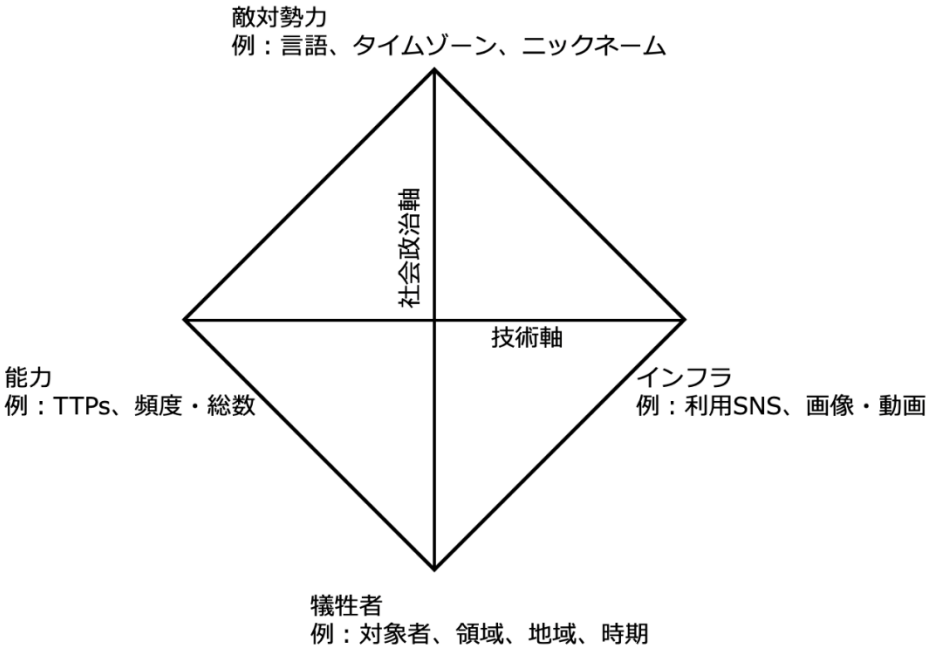
「処理」では、類似した投稿をグループ化するなどして侵入セット（Intrusion Set）を特定する[4]。アクターのデータベース（以下 DB）化と偽情報の選別も行う。類似するキャンペーンのデータがより多くあれば、アクターの人為的ミスの確率は高くなり有効な侵害指標が採取できることもある。また、統計的精度も高まる。しかし、すべてを取得するとデータは膨大になるので、取得データの取捨選択も必要である。さらに、別のキャンペーンでの分析で利用できるように、再利用可能な形に整形して保存する必要がある。

[1] 侵害指標とは、一般にシステム侵害や攻撃を特定・検出するための情報のこと。ここでは、素データに加え、ミームやナラティブを想定する。
[2] クロノロジーとは、事象や出来事を時間的な順序に沿って整理した一覧表とする。
[3] キャンペーンとは、一連の目的を持った組織的な活動や行動のことを指す。
[4] 侵入セットとは、特定の攻撃者グループやキャンペーンによって使用される一連の攻撃パターンや手法を指す。ここでは、キャンペーンに用いられるニュースサイトなども想定する。

ダイヤモンド・モデル分析

ダイヤモンド・モデルを用いたアトリビューションの分析とその例について概説する。ダイヤモンド・モデルは、一般に、2 つの直行軸、すなわち、社会政治軸と技術軸でキャンペーンを特定する分析手法である。詳細は、文献[3]を参照されたい。アトリビューションにおいては、この 2 軸を「犠牲者もしくは敵対勢力」及び「能力もしくはインフラ」として、キャンペーンを分類する(図 1 参照)。

図 1:アトリビューションに適用したダイヤモンド・モデル



出典:文献[3]、p.32 より筆者が翻訳し一部改変。

敵対勢力(Adversary):インフルエンサーの属性に関する情報(アカウント作成時期や利用言語など)を指す。アカウントペルソナ以外に、たとえば、国内外メディア、ジャーナリストや在外大使館などからのメディア発信という形態を取る。

能力(Capability):シャープパワーもしくはソフトパワーなどの戦略的方針や[5]、戦術的手法の組み合わせ、つまり、TTPs(Tactics, Techniques, and Procedures)を指す。クリックベイト、ボット活用、ソックパペットなどトロール、偽情報拡散やドキシングなど、それらの組み合わせ(侵入セット)、アカウントの連携性、投稿の頻度もしくは総量がある。詳しくは、文献[1]、[4]を参照されたい。

インフラ(Infrastructure):攻撃を実施するために敵対勢力が利用する物理的または仮想的なリソースを指す。これには、利用する SNS、ブログサイト、ニュースサイト、メディアの形態などが含まれる。

犠牲者(Victim):攻撃の対象となる個人、性別、人種、民族、組織、地域もしくはイベントを指す。

[5] ソフトパワーは、文化や価値観を通じて影響力を行使する手法を指す。シャープパワーは、情報工作により他国の意思決定を変える戦略的手法を指す。

当該キャンペーンをダイヤモンド・モデルで分析することに加え、予めこの 2 軸で過去のキャンペーンを DB 化しておき、過去のキャンペーンと当該キャンペーンとの比較を行う。ダイアモンドのパターンが一致すればするほど、既知グループへの帰属はより明確になる。その上で、クロノロジーとの比較やその他との整合性を確認し、インサイトを導出し、当該キャンペーンのアトリビューションを定める。

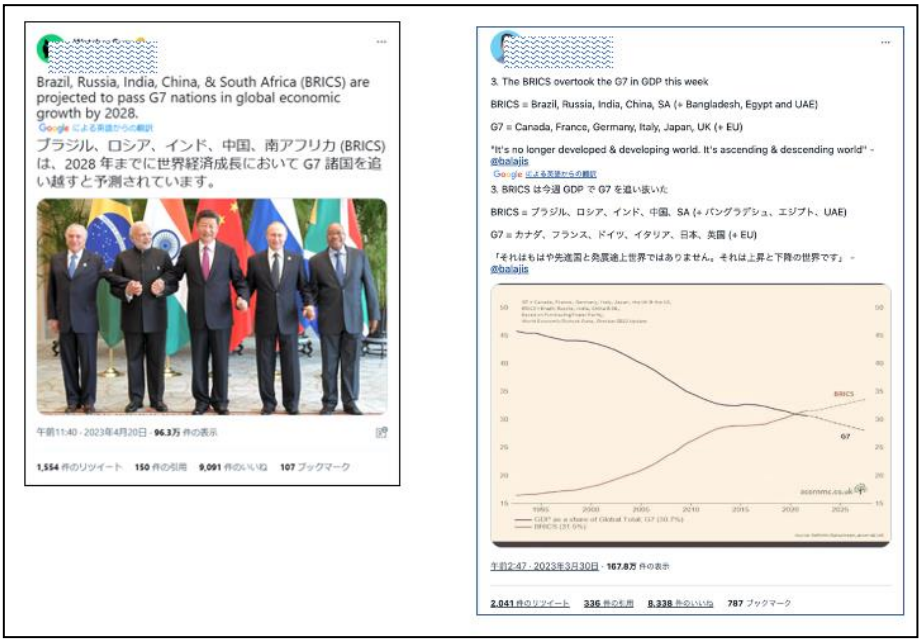
次に、2023 年 G7 外相会議(4/16～18 長野開催)直前に展開された G7 へのキャンペーンを対象として、ダイヤモンド・モデルでの分析例を示す。当該キャンペーンは、G7 外相会議直前で、ピーク時に 1 日 1 万件以上の関連投稿が複数のアカウントから繰り返し行われた。「G7 の経済力は落ち目で、BRICS に抜かれている」という(英語で発信された)ナラティブ、及び図 2 に示す画像などがミームとして X(旧 Twitter)で「協調的に拡散」された[6]。なお、特定ニュースサイトなどへの誘導はあまりないようであった。「敵対勢力」としては、BRICS に関連するアカウントと、ボットと推察されるアカウントが一定数観測されるなど組織的な活動の兆候があった。「犠牲者」は、英語圏で、グローバルサウスの特定地域への発信もあった。

クロノロジーを確認すると、当該 G7 外相会議以降、いくつかのイベントがあった。たとえば、国連の安全保障理事会の選挙が 6 月にあり、グローバルサウスから 2 国(A 国と S 国)が非常任理事国に選出された。さらに、報道によれば、7 月に A 国は「中国と安全保障や国防などで協力強化を図ること」に合意した。S 国においては、中国国家主席が S 国大統領選挙(7 月実施)で再選した大統領に祝電を送ったとの報道もあった。さらに、8 月開催の BRICS の会合で、エジプトなど 6 カ国が新たに加盟することになった。

以上より、当該キャンペーンはシャープパワー寄りの影響力工作であるが、過去の事例から、対外宣伝組織を中心としたキャンペーンであった可能性が推察される。

[6] 「協調的に拡散」とは、複数のアクターが、同一のミームやナラティブを短期間で投稿することを想定する。たとえば、同一内容文の投稿や、同一ハッシュタグのみの投稿などがある。

図 2：2023 年の G7 外相会議前のキャンペーンでの X（旧 Twitter）投稿



出典：筆者の X アカウントにて 2023 年 8 月に採取（一部マスクし自動翻訳を利用した）

アトリビューションの限界

アトリビューションの限界は、情報源をサイバー空間に閉じた場合と、それ以外の情報源を使う場合とでは自ずと変わってくるが、ここでは、サイバー空間に閉じた手法の限界について、概説する。

意図見積であるが、そもそも意図は当人しかわからない。また、状況の変化によりその意図を変えてしまう。よって、意図見積は、ある瞬間の観測による推定である。その上で、情報不足やバイアスがある場合、正しく意図見積ができない。さらに、最終的にアトリビューションの「正解」を得られるとは限らない。過去、諜報活動の暴露により判明したケースもあるが、稀であろう。キャンペーンと陰謀論との関係性もアトリビューションをより困難にさせる。

一般的なサイバー攻撃では、マルウェアや C2 サーバの情報など比較的たくさんの痕跡を侵害指標として入手することができるが、影響力工作のキャンペーンでは入手できるデータに限界がある。また、一部の SNS はデータの取得制限を設けているほか、投稿者が投稿したものを削除するケースもあるので、全てのデータを抽出することが難しくなることがある。膨大な量のデータの中から実際に有効で意味のある情報を見つけ出すのは時間と労力が掛かる。これについては、ツールでカバーできることもある。

意図的に他を模倣することで分析をミスリードする偽旗も想定される[7]。さらに、プロキシ（代替者）を用いたキャンペーンも想定される。攻撃アクターが直接行動するのではなく、間接的に操作や影響を及ぼす第三者や組織を通した活動である。たとえば、ある組織がキャンペーンを仕掛ける

[7] 偽旗とは、敵対的なアクターが、意図的な情報の偽装や、他の TTPs を模倣することで、アトリビューションを失敗させる。

場合、関与を隠蔽しつつキャンペーンを行うことができる対象国内の政治組織を利用することが想定される。資金、ツールや情報などの直接サポートを受けるケース以外に、「理念的な共感」を用いて長期的なキャンペーンを展開するケースが想定される。特に、後者の場合、オリジナルのアクターに辿り着かないこともある。

まとめ

本稿では、IT やデータ分析などの素養のある読者を想定し、筆者らの行っている手法などベースに、サイバー空間での影響力工作におけるアトリビューションについて概観した。インテリジェンス・サイクルのアトリビューションへの適用を示し、G7 外相会議へのキャンペーンをダイヤモンド・モデルで分析した。また、影響力工作におけるアトリビューションの限界についても考察した。紙面の都合もあり不足があることが否めないが、読者に何らかの貢献があれば幸いである。

謝辞

本稿をまとめる機会を与えてくれた市原麻衣子先生に記して感謝致します。

参考文献

[1] 一田和樹、齋藤孝道他『ネット世論操作とデジタル影響工作 ―「見えざる手」を可視化する』（齋藤孝道「第 2 章 デジタル影響工作のプレイブック」、原書房、2023 年」。
[2] 上田篤盛『戦略的インテリジェンス入門』(並木書房、2016 年)。
[3] Timo Steffens, *Attribution of Advanced Persistent Threats: How to Identify the Actors Behind Cyber-Espionage* (Berlin and Heidelberg: Springer Vieweg, 2021).
[4] 齋藤孝道「情報戦における世論誘導工作の片鱗 ―サイバーインフルエンスオペレーションと国内での概況―」『防衛技術ジャーナル』43 巻 1 号(2023 年)、6-14 ページ。

齋藤孝道 プロフィール

明治大学サイバーセキュリティ研究所所長、明治大学理工学部情報科学科教授、レンジフォース株式会社代表取締役、博士(工学)。専門は、サイバーセキュリティ及び、情報セキュリティ技術。特に、Web セキュリティ、ブラウザのトラッキング(ブラウザーフィンガープリント)技術や、サイバー空間での影響力工作を研究テーマとする。著書に、『マスタリング TCP/IP 情報セキュリティ編(第二版)』(オーム社)、『ネット世論操作とデジタル影響工作 ―「見えざる手」を可視化する』(原書房)などがある。