サイバー空間における効果的な抑止メカニズム

熊取谷　行

はじめに

インターネット人口は 2019 年には 41 億人を超え、パソコンのみならずスマートフォン、IOT 機器など、社会のネットワーク化が急速に進展しインターネットへの依存度が増すことに比例して、サイバー攻撃は大きな脅威になっている。

軍事面においても、軍事空間が従来の陸・海・空の空間から、サイバー空間へと拡大するとともに、エストニアやジョージアの事例、あるいはイランに対するスタックスネット・ウイルス攻撃の事例が示すとおり、サイバー攻撃への対処は安全保障上の重要な課題となっている。このため、世界各国は、国家レベルのサイバーセキュリティ政策を推進するとともに、米軍はサイバー・コマンド、中国人民解放軍は戦略支援部隊を新編する等、各国軍隊はサイバー戦専任部隊を設置し、サイバー戦への備えを進めつつある[1]。

抑止メカニズムに目を向けると、冷戦期、米ソ対立構造に一定の安定性を与えた核及び通常戦力によって構成される従来の抑止メカニズムは、敵に受け入れ難いコストを強要する懲罰的抑止が中心であり、その他に敵の特定の行動の利得を拒否する拒否的抑止も存在する[2]。ところが、サイバー空間には陸・海・空といった物理空間とは異なる特徴により、従来の抑止メカニズムが十分に機能しない可能性があるため、サイバー空間における効果的な抑止が模索されているが、必ずしも意見は一致したものではない。特に、サイバー攻撃の報復手段として物理攻撃を用いるかどうかについては意見が分かれており、サイバー攻撃が武力攻撃に相当するか否かの基準となる国際規範の策定に際しても、専門家の間で意見は一致していない現状がある[3]。

---

[1] 山口嘉大「サイバー防衛における官民連携の強化について－エストニア共和国との比較を通じて－」『防衛研究所紀要』第 21 巻第 1 号、2018 年 12 月、163 頁。
[2] 八木直人「抑止概念の再考－新たな脅威様相と「テーラード」抑止－」『海幹校戦略研究』第 2 巻第 2 号、2012 年 12 月、109 頁。
[3] 中谷和弘、河野桂子、黒崎将広『サイバー攻撃の国際法－タリン・マニュアル 2.0 の解説－』信山社、2018 年、77-78 頁。

本稿は、サイバー空間における特徴及び従来の抑止の概念を踏まえた上で、サイバー空間における効果的な抑止メカニズムについて分析、整理することを目的としている。まずは、サイバー空間における抑止を理解するのに必要な、同空間の特徴について述べ、次に、冷戦期の抑止メカニズムとその制約について述べる。そして、サイバー空間において冷戦期の抑止メカニズムが適用できるかどうか述べた後、サイバー空間における効果的な抑止の可能性とその制約について、整理、分析する。

1　サイバー空間の特徴と冷戦期の抑止理論

（1）サイバー空間の特徴

言うまでもなく、サイバー空間は、陸・海・空・宇宙という他の空間が自然空間であるのに対し人工的に作られた空間である。その実態は、パソコン、サーバーといった端末と、それらを接続するネットワークで構成されており、情報の交換、蓄積ができる空間と定義されている[4]。サイバー空間は、このような人工空間であることが影響して、陸・海・空・宇宙という物理空間にはない、いくつかの特徴がある[5]。このうち、本稿では抑止に大きく関係する、「アトリビューション問題」、「攻撃側の優位性」、「効果予測の困難性」、「非対称的なリスク」、「武力攻撃との関係」の五つの特徴に絞って説明する。

第一に、サイバー空間における最大の問題の一つである、アトリビューション問題である。アトリビューションとは、一般的には「所属や属性」という意味だが、サイバー空間においては、サイバー攻撃の主体が誰なのかということを意味する[6]。サイバー空間におけるアトリビューション問題は、インターネット構造、アプリケーションやプログラム設計、攻撃者の社会的属性（特に国家との関係）と多岐にわたるが、大きく分けて技術的なものと社会・政治的なものに区分できる。技術的な問題は、発信元を識別するための IP アドレスを偽装したり、マルウェアによって乗っ取った複数のコンピューターを中継して攻撃を行うボットネットと呼ばれる攻撃手

[4] 持永大、村野正泰、土屋大洋『サイバー空間を支配する者－21 世紀の国家、組織、個人の戦略－』日本経済新聞出版社、2018 年、23 頁。
[5] サイバー空間の特徴は論者によって多少の違いがあるため、本稿では抑止に影響を及ぼす主に攻撃に関する特徴を取り上げた。）
[6] 土屋大洋『サイバーセキュリティと国際政治』千倉書房、2015 年、14 頁。

法などが、攻撃源の特定を困難にしている[7]。社会・政治的な問題は、サイバー攻撃の主体は国家や軍隊に限定されず、個人や犯罪組織など多様であること、また、攻撃者が個人であっても、その背後に国家や組織が存在する場合が挙げられる[8]。攻撃を行ったコンピューターの特定は、技術が進展すれば解決できるかもしれないが、その攻撃を指示した者のアトリビューションを特定することはインテリジェンスの問題となり、技術的な問題よりも困難を極める可能性がある[9]。アトリビューション問題の実事例を示すと、2007 年、エストニアがサーバーやウェブサイトに大規模なサイバー攻撃を受けた際、その多くはロシアからのものであり、攻撃はロシア国内から来ているように見えたが、これが国家が行ったものなのか、国家以外が行ったものなのかは確認できなかった[10]。

第二に、サイバー空間におけるもう一つの大きな問題である、サイバー空間では攻撃側が圧倒的に優位という点である。これは、サイバー空間は侵入や攻撃が極めて容易な空間であり、それらに対する防御が困難という面を持つためである。この性質の要因は複数あるが、根本的には、サイバー空間、とりわけインターネットがその設計上、情報の伝達・拡散を自由かつ容易にすることを目的としているため、リスクマネジメントや安全保障を優先する概念がないことに起因する。また、攻撃側は、自身が攻撃をするための防御側の脆弱性を十分な時間をかけて一つ又は複数見つければよいが、防御側は、常に様々な攻撃に対して防御をしなければならず、わずかな脆弱性を見落としただけで対策が意味をなさなくなる可能性もある[11]。さらに、サイバー攻撃対策は単体のシステムにとどまらず、他のシステムへ被害が拡散する恐れがあるため、ネットワーク全体の横断的な対策が必要となる[12]。よって、少ないコストや資源で実施できる攻撃側に対して、防御側に必要なそのコスト差はあまりに大きい。例えば、1,000 万行のセキュ

---

[7] 川口貴久「サイバー空間における安全保障の現状と課題－サイバー空間の抑止力と日米同盟－」『平成 25 年度外務省外交・安全保障調査研究事業（調査研究事業）「グローバル・コモンズ（サイバー空間、宇宙、北極海）における日米同盟の新しい課題」』、2014 年 3 月、14-15 頁。
[8] 土屋『サイバーセキュリティと国際政治』15-19 頁。
[9] 持永ほか『サイバー空間を支配する者』150 頁。
[10] Martin C. LibiCki, *Cyberdeterrence and Cyberwar*, RAND Corporation, 2009, pp. 2-3.
[11] 川口「サイバー空間における安全保障の現状と課題」16-17 頁。
[12] 持永ほか『サイバー空間を支配する者』148 頁。

リティプログラムに対して、わずか 125 行の強力なマルウェアが作成されることもある[13]。

第三に、サイバー空間においては、攻撃による効果の予測が困難で不確実であり、意図しない結果を招く可能性がある。これは、標的の脆弱性は修正プログラムを適用することにより攻撃を受ける前に是正されるかもしれず、また、ネットワークが予想よりも回復力が高い可能性もあるため、攻撃をしても効果が低い又は効果がない可能性がある[14]。反対に、攻撃の効果がネットワークを通じて広範囲に拡散することで大規模な被害をもたらし、意図しない付随的被害が発生する可能性もある[15]。この実事例として、2010 年にイラン核施設に対して、スタックスネット（Stuxnet）と呼ばれるマルウェアを用いたサイバー攻撃が発生したが、このマルウェアは同施設の遠心分離機のみを標的としていたものの、意図せずインターネット上に流出し、インドネシア、インド、アゼルバイジャン、さらには米国のコンピューターにまで感染が拡大した[16]。

第四に、攻撃に対する非対称的な脆弱性リスクがある。これは、当然のことであるが、コンピューター・ネットワークインフラの環境が進んでいない国は、サイバー攻撃に対する影響が小さくサイバー攻撃を受ける恐れは小さいが、同環境が進んでいる国は、逆にサイバー攻撃に対する影響が大きくサイバー攻撃を受ける恐れが大きいという非対称な脆弱性リスクがあることを意味する。例えば、社会的に ICT が進んでいる米国はサイバー空間への依存度は高いが、米国へサイバー攻撃を仕掛けてくる攻撃国は米国に比べて ICT が進んでいない場合があり、その場合は仮に攻撃を受けた米国はサイバー攻撃で反撃しても、自国が受けた被害と同程度の効果を与えられる可能性は低くなる[17]。

第五に、サイバー空間における攻撃は、それが武力攻撃になり得るかという問題がある。攻撃効果の面だけをみれば、「サイバー攻撃」と呼ばれる

---

[13] Department of Defense, *Defense.gov Deputy Secretary of Defense Speech: Remarks on Cyber at the RSA Conference*, February 15, 2011, archive.defense.gov/speeches/speech.aspx?speechid=1535.
[14] Joseph S. Nye Jr., "Deterrence and Dissuasion in Cyberspace," *International Security* Vol. 41, No. 3, Winter 2016/2017, p. 48.
[15] エリノア・スローン『現代の軍事戦略入門（増補新版）』奥山真司、平山茂敏訳、芙蓉書房、2019 年、305、322 頁。
[16] Symantec, "W32. Stuxnet Dossier," *Symantec Security Response*, November 2010.
[17] Clorinda Trujillo, "The Limits of Cyberspace Deterrence," *Joint Force Quarterly,* Vol. 75, 4th Quarter, 2014, pp. 47-48.

もののほとんどは、他の空間において「攻撃」と呼ばれるものと同等視できるものではなく、武力攻撃になり得るものではない。一方で、電力設備や鉄道管制システム等の重要インフラへのサイバー攻撃では大規模な被害が起こり得るとして、武力攻撃になり得ると論じる専門家もいる[18]。また、サイバー空間を既存の国際法に照らして作成された「サイバー戦に適用される国際法タリン・マニュアル 2.0（Tallinn Manual 2.0 on the International Law Applicable to Cyber Warfare）」[19]（以下「タリン・マニュアル 2.0」という。）では、規則 71（武力攻撃に対する自衛）において「武力攻撃の水準に至るサイバー行動の目標となる国家は、固有の自衛権を行使することができる。サイバー行動が武力攻撃に該当するか否かは、その規模と効果による」とされており、明確には定義されていない。タリン・マニュアル 2.0 策定に当たった専門家の間においても、前述のスタックスネットを用いたサイバー攻撃や国際証券取引所が機能停止、あるいは重要インフラに深刻な影響をもたらすサイバー攻撃が武力攻撃になり得るかどうかについて、意見は一致しなかった[20]。

本稿では、これらの特徴のうち、アトリビューション問題を中心に以降において抑止理論との関連性を分析する。

（2）冷戦期の抑止理論

シェリング（Thomas Schelling）が『軍備と影響力（*Arms and Influence*）』で示した概念は、現在においても抑止概念の基本となっている。シェリングは、抑止の脅威が潜在敵国に理解される場合に限り、抑止の信頼性が保たれると主張し、抑止の脅威と行動が比例均衡していれば、抑止が機能すると分析した[21]。抑止とは戦略的相互作用であり、主体が敵に対して、特定の措置を執るためのコストが増加する可能性を確信させることによって、その措置の採用を防ぐ行為といえる[22]。この抑止の概念は、一般的に懲罰的抑止と拒否的抑止に分類される。懲罰的抑止とは、相手が獲得しようとす

---

18 スローン『現代の軍事戦略入門』319 頁。

19 2017 年に NATO サイバー防衛センター（NATO Cooperative Cyber Defence Centre of Excellence: NATO CCD COE）の専門家らが起草し刊行されたもので、非公式文書ではあるものの、政府機関の要請を受けて作成されたものであるため、一定の権威を有している。「タリン・マニュアル 2.0」の詳細は、中谷ほか『サイバー攻撃の国際法』を参照のこと。

20 中谷ほか『サイバー攻撃の国際法』77-78 頁。

21 Thomas C. Schelling, *Arms and Influence*, Yale University Press, 1966, pp. 141-151.

22 八木「抑止概念の再考」108 頁。

る利益を上回る損失を与えるという報復の脅しによって相手に行動を踏みとどまらせることを意味し、核兵器による報復戦略がその典型的な例である。一方の拒否的抑止とは、相手の目的達成を阻止する可能性を高めて相手に行動は無意味だと悟らせることを意味し、侵略行為の撃退、ミサイル防衛（MD）、重要施設のテロ攻撃からの防護等が該当する[23]。抑止の概念は第二次世界大戦以前にも存在したが、その研究の進展は核兵器の発展と密接に関連している。これは、核兵器はその破壊力と効果的な防衛の困難性に鑑み、抑止によって相手の核兵器使用を防止することが研究の大半となったためである。したがって、抑止といえば、核兵器による報復を示唆しながら相手に核兵器使用を思いとどまらせる懲罰的抑止が一般的となり、拒否的抑止の重要性は大幅に低下した[24]。

懲罰的抑止が成立するためには、被抑止側に攻撃したら容認できないほどの被害を受けると信じさせることが必要となる。ここで重要なのは、このような被害を与えられる能力を抑止側が持っていることを被抑止側が信じて、その被害が容認できないと結論付けさせることであり、これが攻撃しない理由となる[25]。そのためには、アトリビューション、伝達及び信ぴょう性の三つの要件が必要となる[26]。アトリビューションは、サイバー空間におけるものと同じく、攻撃が発生した場合に、それがどこの国家、組織等からのものか特定できることを意味している。冷戦期においては、米ソ二極対立構造であったため対象となる相手が明確であったが、もし、攻撃されても攻撃源が特定できなければ報復攻撃をしようにもできないため、被抑止側に対して攻撃を踏みとどまらせる抑止機能が働かなくなることになる。伝達は、抑止のためには、被抑止側に明確に意図を伝達することで、抑止側が何を思いとどまらせようとしているのか、仮に抑止側の警告を無視して攻撃に踏み切った場合、どのような報復を招くことになるのかということを理解されることを意味している。信ぴょう性は、抑止が成立するか否かは、最終的には被抑止側の意思決定であるため、被抑止側が攻撃に踏み切った場合には、抑止側が報復攻撃をする意思と能力を有することを

---

[23] 福田毅「抑止理論における『第 4 の波』と冷戦後の米国の抑止政策」『日本国際政治学会 2012 年度研究大会部会 13「地域抑止」の現状と課題』、2012 年 10 月、1 頁。

[24] 川口「サイバー空間における安全保障の現状と課題」13 頁。

[25] Patrick M. Morgan, “Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm,” National Research Council of The National Academies, ed., *Proceedings of a Workshop on Deterring Cyberattacks*, The National Academies Press, 2010, p. 61.

[26] Trujillo, “The Limits of Cyberspace Deterrence,” p. 45.

被抑止側が信じることを意味している。理論的には、これら三つの要件が満たされる場合、つまり、抑止側が対象である被抑止側を特定し、同相手に正しく意図を伝達し、同相手の行動を踏みとどまらせるに足る信ぴょう性を持っていれば、被抑止側は、攻撃に踏み切ることによって得られる利益と報復攻撃で負わされるコストを比較し、後者が上回る場合には、攻撃を踏みとどまることになる[27]。よって、懲罰的抑止は、相手に関する信頼性の高い知識、信ぴょう性の高い軍事力、そして相手との一定の共有された意思疎通が必要になると言える[28]。

他方、このような抑止メカニズムには制約もある。冷戦期に発展した抑止理論は、結果的に米ソ超大国間での核戦争が起こらなかったため成功したと考えられている。しかし、それは核戦争が「起こらなかった」という事象によって判断されているものであり、果たして抑止が本当に機能したのかどうかを評価することは非常に困難である[29]。この点を、元米国家安全保障担当補佐官キッシンジャー（Henry Kissinger）は、「抑止とは、実際に『起こらなかった』ことから否定的な形でのみ検証出来るだけであり、また、何かが起らなかったことを示すことは不可能である」と指摘している[30]。また、抑止には技術、軍事、政治、外交等、いくつもの変数があり、これらが時間とともに変化するため、長期にわたって同じ抑止メカニズムが機能することはほぼ不可能である。さらに、抑止は特定の相手には有効であっても、その他の相手には有効とは限らず、潜在的な攻撃者の全てを抑止できるとは限らない[31]。抑止は、このような制約と予期せぬ事態によって脆く崩れる危険性をはらんでいることを認識する必要がある。

2　サイバー空間における抑止の問題

サイバー空間における抑止は、前節で述べた攻撃側の優位性が冷戦期の核抑止も同様に攻撃側が優位であったという特徴による類推もあり、サイ

---

27 栗田真広「サイバー攻撃に対する『抑止』の現状－米国の安全保障政策の事例から－」『科学技術に関する調査プロジェクト調査報告書「情報通信をめぐる諸課題（科学技術に関する調査プロジェクト 2014）」』、2015 年 3 月、160 頁。
28 アントゥリオ・エチェヴァリア『軍事戦略入門』前田祐司訳、創元社、2019 年、85 頁。
29 スローン『現代の軍事戦略入門』130 頁。
30 ヘンリー・A・キッシンジャー『外交（下）』岡崎久彦監訳、日本経済新聞社、1996 年、231 頁。
31 エチェヴァリア『軍事戦略入門』74-75 頁。

バー空間における懲罰的抑止の議論へとつながった。しかし、同じく前節で述べたアトリビューション問題等のその他のサイバー空間の特徴によって、サイバー空間の抑止についての政策・研究の多くは、冷戦期の懲罰的抑止モデルはサイバー空間では機能しないという見方を示した[32]。

オバマ政権で米国防副長官を務めたリン（William Lynn）は「一度のクリックは 0.3 秒で地球を 2 周する。その一方で、攻撃源を特定するのには数か月を要する。ほぼリアルタイムでサイバー攻撃者を特定しなければ、我々の抑止プログラムは破綻する。ミサイルは『返信先』を明らかにしてやってくるが、サイバー攻撃の多くはそうではない。このような理由で、既存の抑止モデルは、サイバー空間では全く当てはまらない」と言っている[33]。また、クラーク（Richard Clarke）らによる『サイバーウォー（*Cyber War*）』では、サイバー戦争を「損害ないし破壊を引き起こす目的のために他の国家のコンピューターないしネットワークに侵入する国民国家による行動」と定義した上で、核兵器と比較しながらサイバー空間における抑止について、「核戦略の概念の中では、抑止理論はおそらくサイバー戦争には最も適用しにくいものである。実際、サイバー空間における抑止力は、ハーマン・カーンや 1960 年代の戦略家の作品の中での抑止力とは、全く異なる意味を持つことになるだろう」と論じている[34]。

このような、サイバー空間において冷戦期の抑止メカニズムをそのまま適用した場合には、十分に機能しない可能性が高いとする主張の理由は、次のとおりサイバー空間では懲罰的抑止と拒否的抑止の両方に問題が存在するためである。

（1）懲罰的抑止の問題

懲罰的抑止は前節で述べたとおり、抑止側は被抑止側を特定し、その相手に対して攻撃を行った場合には報復するという脅しを伝達し、それを信頼させなければならないというアトリビューション、伝達、信ぴょう性の要件が必要だが、サイバー空間においては次の三つの問題がある。

第一に、第三者のコンピューターを中継する等の、攻撃源を偽装することの多いサイバー攻撃のアトリビューションを特定するには、高度な技術

---

[32] 栗田「サイバー攻撃に対する『抑止』の現状」161 頁。
[33] Kevin P. Chilton and William J. Lynn III, *2010 Cyberspace Symposium: Keynote – DoD Perspective*, May 26, 2010, archive.defense.gov/speeches/speech.aspx?speechid=1477, accessed July 12, 2020.
[34] Richard A. Clarke and Robert K. Knake, *Cyber War*, Harper Collin Publishers, 2010, pp. 92.

と多大な時間・労力を要するため、攻撃者の特定は極めて難しい。仮に抑止側が攻撃源を特定した場合であっても、攻撃源が本来の攻撃者なのか、中継として利用された第三者なのか確実な判断ができない場合は、報復攻撃を躊躇せざるを得なくなる[35]。また、サイバー攻撃は、国家に限らず犯罪組織、個人といった主体が多様であるため、対象となる相手が単一でなく、何を抑止の対象とするのかという面も懲罰的抑止を困難にしている[36]。

第二に、アトリビューションが不確実であることは、正しい相手に伝達することが困難であり、間違った相手に伝達した場合は、新たな紛争の火種になる可能性がある[37]。また、アトリビューションが特定できた場合であっても、被抑止側にサイバー攻撃によって標的を危険に晒すと伝達すると、被抑止側は標的をネットワークから切り離す等の防御手段を講じる可能性があり、報復攻撃が役に立たなくなる可能性がある[38]。

第三に、核攻撃の破壊力はよく知られており、核戦略における懲罰的抑止の信ぴょう性はこの十分な確実性から生まれた。一方、サイバー攻撃の多くは、攻撃効果が限定的又は情報の窃取やネットワークへの侵入といった諜報的なものであり、核兵器ほど相手の信ぴょう性を得られるような説得力のあるものではない。仮に強力なサイバー兵器を持っていた場合は、被抑止側から十分な信ぴょう性を得るために、その能力や攻撃方法等の詳細を伝達する必要があるが、詳細を伝達した場合は相手が攻撃に対抗する防御手段を講じる可能性が生じてしまう[39]。

サイバー空間には、このような懲罰的抑止の要件に係るものの他にも問題が存在する。抑止を確立する上で報復攻撃効果の予測は非常に重要であるが、報復のサイバー攻撃の対象である標的の脆弱性は、相手に発見されればたちまち修正されて攻撃が役に立たなくなるという不確実性がある[40]。また、攻撃による被害が相手国のみならず、意図せず第三国や自国にまで及ぶかもしれないという、効果の予測が困難という問題もある[41]。さらに、核攻撃は、攻撃と報復の応酬が一巡するころには双方ともに反撃できる能力が残っていない程の被害を受けている可能性が高いが、サイバー攻撃で

---

35 山口「サイバー防衛における官民連携の強化について」165 頁。
36 栗田「サイバー攻撃に対する『抑止』の現状」161-163 頁。
37 Mariarosaria Taddeo, “The Limits of Deterrence Theory in Cyberspace,” *Philosophy & Technology,* Vol. 31, No. 3, September 2018, p. 352.
38 Trujillo, “The Limits of Cyberspace Deterrence,” p. 48.
39 Clarke and Knake, *Cyber War*, pp. 92-95.
40 LibiCki, *Cyberdeterrence and Cyberwar*, p. 55.
41 スローン『現代の軍事戦略入門』322-323 頁。

は核攻撃ほどの破壊力はなく、攻撃と報復が繰り返し行われる可能性がある[42]。その場合は、米国のような ICT 先進国は、サイバー空間における非対称的な脆弱性リスクのため、報復の相手よりもサイバー能力が優位であるにも関わらず、相手よりも大きな被害を受けるリスクが高くなる[43]。

（2）拒否的抑止の問題

このようなサイバー空間における懲罰的抑止の問題から、冷戦期には重要性が低下した拒否的抑止が注目され議論されるようになり、例えば、リビッキー（Martin LibiCki）は懲罰よりも防衛（拒否的抑止）がより大きな役割を果たしていると主張した[44]。また、2010 年にリン米国防副長官が『フォーリン・アフェアーズ・リポート（*Foreign Affairs Report*）』誌に寄せた論説においては、「抑止を機能させるには、報復措置でコストを強いるのではなく、攻撃者のあらゆる利益を否定することを重視する必要がある」として、サイバー空間における拒否的抑止の重要性について主張している[45]。

しかしながら、前節で述べたとおり、そもそもサイバー空間においては攻撃側が圧倒的に優位であり、防御は困難という特徴がある。仮に防御が一時的に成功しても、サイバー空間のアトリビューション問題により攻撃源の特定は困難なため、サイバー攻撃は継続する場合があり、攻撃者の攻撃意図を抑止することができない可能性がある[46]。また、ソフトウェアのセキュリティ上の脆弱性で一般的に知られていないものは「ゼロデイ脆弱性」と呼ばれるが、このゼロデイ脆弱性を完全に除去して完全な防御を実現しようとした場合、莫大なコストが必要となる。一方で、攻撃者は、安価に脆弱性を選択して攻撃を実施することができるため、拒否的抑止を機能させることが困難となる[47]。

---

42 LibiCki, *Cyberdeterrence and Cyberwar*, pp. 30-31.
43 Ibid., p. 32.
44 Ibid., p. 7.
45 ウィリアム・J・リン三世「ペンタゴンの新サイバー戦略―なぜアメリカはサイバー軍を立ち上げたか」『フォーリン・アフェアーズ・リポート』、2010 年 10 月号、18-27 頁。
46 Taddeo, "The Limits of Deterrence Theory in Cyberspace," pp. 346-347.
47 山口「サイバー防衛における官民連携の強化について」165-166 頁。

3　サイバー空間における抑止の可能性と制約

（1）サイバー空間における効果的な抑止理論の模索

前節で述べたとおり、サイバー空間において冷戦期の抑止メカニズムを適用することは、様々な問題があるため困難と考えられてきたが、専門家において冷戦期の抑止メカニズムの適用の可能性も含めて、サイバー空間における効果的な抑止が模索されている。

懲罰的抑止の最も大きな問題であるアトリビューション問題については、ナイ（Joseph Nye）は、「アトリビューションは程度の問題である。法廷で立証されるような質の高いアトリビューションを得ることは難しいが、抑止が可能になるアトリビューションが存在することは多い。また、迅速で質の高いアトリビューションは困難でコストがかかるが、不可能ではない。不完全なアトリビューションがあるからといって、懲罰的抑止ができないわけではない」と、現状においてもある程度のアトリビューションは可能と主張している[48]。また、2011 年 11 月に米国防省が議会に提出した「サイバースペース政策報告（Department of Defense Cyberspace Policy Report）」の中では、攻撃の発信源を特定する能力の向上に向けた努力と、その成果を強調しており、詳細は明らかではないがアトリビューション問題解決に向けた技術的な能力が向上していることが推察できる[49]。技術的な能力が向上しても、サイバー攻撃の主体が国家なのか、背後に国家や組織が存在するかどうかという社会・政治的な問題は残るが、この問題については、「誰が攻撃を行ったのか」ではなく、「誰がその行為に対して責任を取るのか」を追及することによって、アトリビューション問題を解決していこうとする新しいアプローチの主張がある[50]。

拒否的抑止の問題については、全ての脆弱性への対応は難しいものの、国家レベルと比較すると技術レベルが劣ると思われる非国家や犯罪組織を対象とする場合は、縦深的な防御装置の設置等による強力な防御によって、拒否的抑止が機能するという主張がある[51]。また、米国務省国際安全保障諮問委員会の「国際的なサイバー安定性の枠組みに関する報告」（2014 年 7

---

48 Nye, “Deterrence and Dissuasion in Cyberspace,” pp. 51-52.
49 Department of Defense, *Department of Defense Cyberspace Policy Report: A Report to Congress Pursuant to the National Defense Authorization Act for Fiscal Year, Section 934,* November 2011, p. 4.
50 Jason Healey, “Beyond Attribution: Seeking National Responsibility for Cyber Attacks,” *Atlantic Council Issue Brief*, January 2012, pp. 1-7.
51 Dorothy E. Denning, “Rethinking the Cyber Domain and Deterrence,” *Joint Force Quarterly,* Vol. 77, 2nd Quarter, 2015, p. 12.

月）の中で、サイバー空間の抑止力は拒否的抑止、懲罰的抑止の他に、攻撃対象のアーキテクチャがレジリエントであると思わせること（レジリエンスによる抑止）から構成されるとし、レジリエンスによる抑止という概念が追加されている[52]。レジリエンスとは、米国防省サイバー戦略（2015 年 4 月）において、その目的は国防省のネットワークが破壊的又は妨害的なサイバー攻撃を受けたとしても機能を継続させるものとされている[53]。

このように、困難とされてきた懲罰的抑止と拒否的抑止のサイバー空間への適用については、様々な考え方や手段によって解決の可能性を見出そうと模索されているが、この他にも伝達や信ぴょう性といった多くの問題が残っている。そこで提唱されているのが、懲罰的抑止、拒否的抑止を含めた複数の手段を組み合わせる考え方である。

ナイは、サイバー空間で抑止が機能するかどうかは、「どのように」、「誰が」、「何をするか」であり、アトリビューションの困難さや主体の多様性があるからといって、サイバー空間での抑止が不可能になるわけではないとして、従来の「懲罰（Punishment）」、「拒否（Denial）」に加えて「絡み合い（Entanglement）」、「規範（Norms）」の 4 つの抑止機能を提唱している[54]。「懲罰」は、懲罰的抑止が冷戦期ほどの役割を果たすことはないとしながらも、抑止の重要な部分であることは変わりないとして、報復攻撃の手段をサイバー攻撃に限定せず、外交、経済、物理という他の手段の使用を肯定している。「拒否」は、攻撃者のリソースと時間は限られているため、優れたサイバー防御は攻撃者のコストを押し上げ、攻撃を抑止できる効果があるものとしている。「絡み合い」とは、様々な相互依存関係を指しており、この依存関係によって、たとえ攻撃に対する報復の恐れがなかったとしても関係悪化を恐れて、現状維持の継続にメリットがある場合は攻撃を行わないとしている。「規範」とは、規範とタブーを指しており、攻撃者が攻撃によって得られるメリットよりも規範やタブーを破ることで著しく信頼や評判が低下する等、コストがそれを上回る場合があるとしている。そして、これら「懲罰」、「拒否」、「絡み合い」、「規範」の 4 つの機能は、いずれも完璧なものではないが、表のように、これを「どのように」、「誰が」、「何をするか」によって組み合わせることで、効果的な抑止が機能すると主張している[55]。また、ナイが主張するような抑止の対象によってその方法

[52] United States Department of State, *Report on A Framework for International Cyber Stability*, July 2014, p. 11.
[53] Department of Defense, *The DoD Cyber Strategy*, April 2015, p. 11.
[54] Nye, "Deterrence and Dissuasion in Cyberspace," pp. 54-69.
[55] Ibid.

を変える考え方は、米国が 2006 年の「4 年毎の国防見直し（Quadrennial Defense Review: QDR）」で示した、抑止をオーダーメイドなアプローチへ転換する「テーラード抑止（tailored deterrence）」と呼ばれる抑止概念と共通するものがある[56]。クレイマー（Franklin Kramer）らは、2013 年発表の報告書でサイバー空間での安全保障政策において、テーラード抑止を採用することの重要性を提唱している[57]。

表　「懲罰」「拒否」「絡み合い」「規範」を組み合わせるメカニズム

どのように (HOW)	懲　罰 (Punishment)	拒　否 (Denial)	絡み合い (Entanglement)	規　範 (Norms)
誰を (WHO)	国家と非国家の両方	小規模国家や非国家	中国などの大国	主要国家、非国家、犯罪者
何を (WHAT)	主要な武力行使：武力紛争未満のレベルの行動	犯罪やハッキング（ただし、先進国に対しては不完全）	主要な武力行使：武力紛争未満のレベルの行動	武力行使レベル（武力紛争法で対応）、民間人（規制で対応）、犯罪（規範で対応）

（出所）Joseph S. Nye Jr., “Deterrence and Dissuasion in Cyberspace,” *International Security,* Vol. 41, No. 3, Winter 2016/2017, p. 69.

モーガン（Patrick Morgan）は、ナイと同じように攻撃に対して報復によって対応することは抑止がある程度機能するとして、慎重かつ適切に行うものとした上で、経済的、軍事的、外交的といった物理攻撃を含むサイバー攻撃以外の様々な対応が必要と述べている。また、懲罰的抑止は機能するかどうか不確実であるため、拒否的抑止の強化が一層重要として、攻撃を迅速に検知・対応する能力や、ネットワークやシステムに侵入されても防御できる能力が重要であると主張している。さらに、紛争や武器に関する多国間の安全保障の取り組みのように、サイバー空間における集団的な軍備管理を積極的に推進し、サイバー空間の再編と同空間の規範を監督するための新たな組織作りを提唱している[58]。

[56] Department of Defense, *Quadrennial Defense Review Report*, February 2006, p. 49.
[57] Franklin D. Kramer and Melanie J. Teplinsky, “Cybersecurity and Tailored Deterrence,” *Atlantic Council Issue Brief*, December 2013, pp. 2-3.
[58] Morgan, “Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm,” pp. 75-76.

第 1 節で述べたように、抑止は機能しているかどうか評価することは非常に困難であるため、これらの主張が適切かどうか評価することは難しい。しかし、少なくとも複数の専門家等からこのような複数の手段を組み合わせる考え方が提唱されていることは、根本的な抑止問題の解決には至らないものの、これらは抑止効果が期待できる考え方の趨勢と言ってよいだろう。

（2）サイバー空間における抑止の制約

提唱されているサイバー空間における効果的な抑止の可能性のある主張には問題も残っている。サイバー空間における抑止の議論において、主張が大きく分かれているのが、報復攻撃の手段に物理攻撃を用いるかどうかという点である。

前述のとおり、サイバー攻撃による報復攻撃は攻撃効果が予測困難であり、また、被抑止側から信ぴょう性を得られない可能性があるという問題がある。このため、報復攻撃の手段をサイバー攻撃に限定せず物理攻撃を用いることで、確実性及び信ぴょう性を確保して抑止を機能させようとするのが、報復手段に物理攻撃を用いることを主張する理由である。これに対し、報復手段に物理攻撃を用いることに否定的な主張は、報復攻撃（物理攻撃）が受ける攻撃（サイバー攻撃）に対して比例したものではなく、攻撃者は報復に応じて物理攻撃、場合によっては核攻撃にまでエスカレーションを招くリスクがあるため、物理攻撃を用いるべきではないとするものである。また、サイバー攻撃に対してサイバー攻撃により報復することは、サイバー空間から物理空間への敷居を踏み越えないという、これ以上のエスカレーションを回避するシグナルを送ることでもあるとも述べている[59]。リビッキーは、「もし米国が物理手段を使ってエスカレートすれば、他国をサイバー空間に留めることで得られるかもしれない利点を失い、なぜ最初に暴力を使ったのか説明しなければならない（サイバー攻撃で死傷者を出さない限り）」と、サイバー攻撃に対する過剰な報復はエスカレーションを招くと警鐘を鳴らしている[60]。また、タデオ（Mariarosaria Taddeo）は、報復手段に物理攻撃を用いることは、報復措置の均衡性、比例性が成り立たず、物理攻撃による報復を正当化してエスカレーションを招く危険性があると主張している[61]。

---

[59] 栗田「サイバー攻撃に対する『抑止』の現状」164-165 頁。
[60] LibiCki, “Cyberdeterrence and Cyberwar,” pp. 70-71.
[61] Taddeo, “The Limits of Deterrence Theory in Cyberspace,” p. 350.

このようなエスカレーションを回避するためには、「対象」と「手段」を考慮したエスカレーション・コントロールが求められるが、現状では、主要国間でいわゆるエスカレーション･ラダーの共通認識が存在しない。よって、たとえ報復が攻撃に対して比例したものであっても、攻撃者は報復が自分の認識の閾値を超えた場合は攻撃をエスカレートさせる可能性があり、低烈度で発生したサイバー攻撃が急速に深刻なサイバー攻撃の応酬、最終的には物理的な戦争へと発展する危険性をはらんでいる[62]。

報復手段に物理攻撃を用いることはエスカレーション問題の他にも、このような報復が国際社会の理解を得られるかどうかは定かではないという問題もある[63]。前述のとおり、タリン・マニュアル 2.0 規則 71（武力攻撃に対する自衛）においても明確な基準は示されておらず、また、どのようなサイバー攻撃が武力攻撃になり得るか、専門家の間でも意見は一致していない。よって、もし、報復手段に物理攻撃を用いる場合は、それによって得られる利益が国際的な非難を受けることによって負うコストを上回る場合、又は、受けたサイバー攻撃が武力攻撃に相当すると国際的な理解を十分に得られるほどの深刻な被害が生じた場合に限定されると考えられる。

さらに、報復手段としてサイバー攻撃をする場合であっても、前述のサイバー攻撃による効果の予測が困難で不確実な問題のため、そのリスクを回避するためには、攻撃による被害範囲の予測が十分に可能な方法又は、被害範囲が広がっても大きな影響を及ぼさない場合に限定する必要がある。ただし､サイバー攻撃による効果の予測が困難で不確実な問題は､サイバー攻撃による報復の程度を適切にコントロールすることを困難にしており、例えば、報復を均衡のとれた措置にしたつもりが、予期せぬ大きな攻撃効果が生じてエスカレーションを招くリスクも抱えている[64]。

おわりに

冷戦期の抑止は、米ソ二極対立構造において核戦争を防ぐことが至上命題であったため、単一の主体に対して絶対的メカニズムが存在し得た。ところが、サイバー空間における抑止は、その空間の特徴、主体の多様性といった要因のため、現在のところ絶対的な抑止メカニズムは存在しない。懲罰的抑止、拒否的抑止の他、多様なアプローチを抑止対象に合わせて組

62 栗田「サイバー攻撃に対する『抑止』の現状」165 頁。
63 同上、164-165 頁。
64 福富「サイバー対抗措置の可能性と限界」45 頁。

み合わせるといった効果が期待できる考え方が主張される一方、サイバー攻撃が武力攻撃に相当するか否かという問題は残ったままである。このような問題を解決するため、国連では、安全保障問題を取り扱う第一委員会の下に、「国際安全保障の文脈における情報通信分野の発展に関する政府専門家グループ（Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security: GGE）」を設け、サイバー空間における安全保障に関する議論を行ってきた。しかし、第 5 会期（2016～2017 年）会合では、国際法の枠組みの中で規範を作るべく米国主導で始めた議論をまとめた最終報告を、西側主導の規範作りに警戒感を示した中国やロシアが受け入れず、報告書の作成には至らなかった。中国とロシアは、国家が国内の情報を統制する権利と責任を持つべきという立場から、サイバー空間には既存の国際法を適用するのではなく新たなルールが必要として、米国主導の報告書とは異なる「情報セキュリティのための国際行動規範（案）」を国連総会に共同提案を行っており、各国間の足並みは揃っていない[65]。このような中、米国トランプ（Donald Trump）政権下では、2018 年 9 月に発表した「国家サイバー戦略（National Cyber Strategy）」において、外国からのサイバー攻撃に対して積極的に攻撃的手段をとることも辞さない方針を表明するとともに、ロシア、中国、イラン、北朝鮮を、米国及び同盟国の経済、民主主義を損ない、知的財産を盗み、民主的プロセスを妨げる敵対国家と見なすなど、一層、攻撃色の強い政策を表明している[66]。

しかし、繰り返しになるが、具体的にどのようなサイバー攻撃が武力攻撃に相当するかどうかは、国際的な合意が得られていない。例えば、2019 年 4 月 19 日に行われた日米安全保障協議委員会（SCC）において、「いかなる場合にサイバー攻撃が日米安全保障条約第 5 条の下での武力攻撃を構成するかは、他の脅威の場合と同様に、日米間の緊密な協議を通じて個別具体的に判断されることを確認した」と判断基準は示されていない[67]。このため、米国はサイバー攻撃に対して物理的な軍事対応を取り得ることを機会ある毎に繰り返し述べており、今後形成される国際的規範の策定の方向

---

65 山崎治「自衛隊、米国軍等のサイバー攻撃対処能力の強化」『レファレンス』832 号、2020 年 5 月、15-16 頁。

66 White House, *National Cyber Strategy of the United States of America*, September 2018, p. 2.

67 「日米安全保障協議会「2＋2」共同発表（仮訳）」外務省、2019 年 4 月 19 日、www.mofa.go.jp/mofaj/files/000470737.pdf。

性に影響力を行使しようと注力している[68]。このような主要国の国際的な規範策定への積極的な関与の姿勢は、サイバー空間における抑止メカニズムが十分ではなく、抑止として取り得る手段の主導権争いと考えられる。

前述のとおり、抑止は証明することが非常に困難であり、また、技術や経済状況によって変化する。したがって、本稿で述べた抑止手段が適切とは限らず、また、仮に同抑止手段が適切であったとしてもいつまでも効果があるものとは限らない。サイバー空間における抑止メカニズムは今も模索中の現在進行形の問題であり、今後も注視していく必要がある。

[68] 栗田「サイバー攻撃に対する『抑止』の現状」179 頁。