The Attribution Threshold: From Technical Confidence to Lawful Response in Cyber-Economic Warfare

Supplementary Paper 6

Christopher I. V. Farmer

Publication edition, August 2026

Contents

Abstract....4
Table of Authorities....6
Cases....6
Legislation....7
Treaties and International Instruments....8
Standards, Government and Technical Materials....8
Corporate, Technical and Contemporary Materials....12
1. The Missing Law of Decision....12
2. Attribution Is Three Questions, Not One....14
2.1 Technical Attribution....14
2.2 Legal Attribution....15
2.3 Decisional Attribution....16
3. The Evidential Object....17
3.1 A Claim Graph, Not a Dossier Label....17
3.2 Direct Fact, Inference and Legal Conclusion....18
3.3 Contrary Material and Live Hypotheses....19
3.4 Preservation, Correction and the Evidential Clock....19
3.5 Numerical Confidence and Calibration....20
4. The Response-Specific Thresholds....21
4.1 Defensive Preservation and Containment....21
4.2 Confidential Diplomatic and Provider Engagement....22
4.3 Public Attribution....23
4.4 Sanctions and Designation....24
4.5 Civil Restraint, Disclosure and Asset Measures....25
4.6 Criminal Investigation, Charge and Conviction....25
4.7 Countermeasures....26
4.8 Force in Self-Defence....27
5. Attribution to States, Auxiliaries and Territorial Hosts....29
5.1 Organs and Empowered Entities....29
5.2 Instructions, Direction and Control....30
5.3 Acknowledgment, Adoption and Assistance....30
5.4 Territorial Knowledge and the Attribution of Omissions....31
5.5 PMSCs, Contractors and Autonomous Agents....31
6. False Flags, Campaigns and the Designed Appearance of Another....32
6.1 False Flags Are Operations, Not Magic....32
6.2 Campaign Aggregation....33
6.3 Four Public Attribution Records....33
7. Intelligence, Closed Material and Allied Confidence....34
7.1 Intelligence May Decide Without Becoming Trial Evidence....34
7.2 Closed Material Requires Law....35
7.3 Allied Intelligence and the Risk of Circular Confidence....36
7.4 The Public Case Without Source Sacrifice....36
8. Fragmentation, Adverse Inference and the Controller’s Evidential Burden....37
8.1 Designed Opacity....37
8.2 No Transfer of Ignorance to the Territorial State....38
8.3 The Manifest as Evidence, Not Immunity....38
9. Error, Correction and Compensation....38

9.1 Error Is a Foreseeable State, Not an Institutional Scandal....38
9.2 The Correction Duty....39
9.3 Remedy and Compensation....39
10. Operational Tests....40
10.1 Destructive Access to NHS Data....40
10.2 Winter Energy Pre-positioning....41
10.3 Theft from an Insurer....41
10.4 The Foreign-Paid PMSC....42
10.5 The Fragmented AI System....42
11. The Attribution Decision Certificate....43
11.1 Content....43
11.2 Threshold Matrix....43
11.3 The Two Error Axes....44
11.4 Review Does Not Repeat the First Decision....45
11.5 Ownership and Institutional Dissent....45
12. Conclusion....46
Bibliography....47
Books, Chapters and Articles....47

Abstract

Cyber-economic warfare presents government with an unpleasant choice which the language of confidence has too often concealed. If the State waits for evidence fit to convict a named operator, the hospital may lose its records, the water utility its control system and the winter grid its margin of safety. If it treats technical suspicion as proof of State responsibility, it may punish the wrong person, strike the wrong object or convert a remediable intrusion into an international crisis. The defect is not that attribution is impossible. It is that law has failed to state, in one place, what must be attributed, to whom, upon what material and with what degree of assurance before each form of response becomes lawful.

This paper supplies that missing law of decision. Technical attribution identifies infrastructure, tooling, access, operator and campaign. Legal attribution asks whether proved conduct is that of a natural person, company, State organ, empowered entity, instructed or controlled auxiliary, adopting State or assisting State. Decisional attribution asks whether the evidence is sufficient for the particular act now proposed. Those questions overlap, but they are not interchangeable. A malware family is not a State; nationality is not control; presence on territory is not knowledge; and a government's public assessment is not a criminal conviction.

The paper establishes a response ladder. Reversible preservation and defensive containment may proceed upon recorded and particularised reasonable grounds, provided that the measure is narrow, short and independently reviewable. Confidential diplomacy may rest upon a candid intelligence assessment which identifies gaps. Public attribution requires corroborated, source-conscious confidence sufficient to justify the accusation to allies, the public and the accused. Sanctions and civil restraint require the statutory or judicial threshold governing the interference and a rational connection between person, conduct and purpose. Charge requires a realistic prospect of conviction; conviction requires the tribunal to be sure. Countermeasures require a sufficiently established prior internationally wrongful act attributable to the State against which they are directed. Force requires a high degree of confidence in the facts necessary to establish an actual or imminent armed attack, necessity, a lawful target and proportionality; uncertainty concerning a target cannot be cured by certainty concerning the danger in general.

Maximum reasonably credible foreseeable harm affects urgency, preservation, precaution and the acceptable cost of temporary defensive error. It does not lower the proof necessary for irreversible punishment or force. The stronger answer to speed is not evidential indulgence but staged authority: preserve first, test hypotheses, contain the smallest dependency, record dissent, demand correction through every recipient and escalate only when the proposition required by the next legal gateway is proved.

Classified intelligence and allied reporting may lawfully inform executive action, but secrecy is not weight. The decision record must distinguish direct fact, technical inference, organisational inference, legal conclusion and policy judgement; identify independence, provenance, deception opportunity and material alternatives; and state what would change the conclusion. Closed procedures, special advocates and public-interest immunity may protect sources only within their lawful fields. They cannot create a secret common law of punishment.

Where a controller deliberately shards an operation, withholds the manifest or destroys the only record capable of connecting payer, operator and effect, the law may draw a bounded adverse inference against that controller. It may not transfer the controller's obstruction to an innocent territorial State which saw only one fragment. Designed opacity changes the evaluation of evidence against its designer; it does not abolish the substantive rules of attribution.

The resulting Attribution Decision Certificate forms the hinge between the MOSAIC evidence architecture and the United Kingdom Economic Warfare Prevention and Response Act proposed in the preceding papers. It records the proposition proved, the applicable threshold, the supporting and contrary material, the decision owner, the measure selected, its maximum duration, the reviewer and the correction route. Firm response and disciplined proof are not opponents. In a domain built upon disguise, speed without a record is rashness, while a standard which can be satisfied only after the threatened harm is complete is not caution but surrender.

Keywords: cyber attribution; economic warfare; evidence; intelligence; sanctions; countermeasures; self-defence; State responsibility; false flags; closed material; adverse inference; decision standards

Table of Authorities

Cases

A and Others v United Kingdom (2009) 49 EHRR 29 35

Ahmed v HM Treasury [2010] UKSC 2, [2010] 2 AC 534 24

Air Services Agreement (United States of America v France) (Award) (1978) 18 RIAA 417 26

Al Rawi v Security Service [2011] UKSC 34, [2012] 1 AC 531 35

American Cyanamid Co v Ethicon Ltd [1975] AC 396 (HL) 406–09 13

Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro) [2007] ICJ Rep 43 15

Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda) (Judgment) [2005] ICJ Rep 168 28

Armory v Delamirie (1722) 1 Strange 505, 93 ER 664 37

Bank Mellat v HM Treasury (No 1) [2013] UKSC 38, [2014] AC 700 35

Bank Mellat v HM Treasury (No 2) [2013] UKSC 39, [2014] AC 700 13

Bankers Trust Co v Shapira [1980] 1 WLR 1274 (CA) 1282 25

Big Brother Watch and Others v United Kingdom (2022) 74 EHRR 17 19

Conway v Rimmer [1968] AC 910 (HL) 952–53 35

Corfu Channel (United Kingdom v Albania) (Merits) [1949] ICJ Rep 4 18

Council of Civil Service Unions v Minister for the Civil Service [1985] AC 374 (HL) 410–11 39

Difference Relating to Immunity from Legal Process of a Special Rapporteur of the Commission on Human Rights (Advisory Opinion) [1999] ICJ Rep 62 29

Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Kärntner Landesregierung (Joined Cases C-293/12 and C-594/12) EU:C:2014:238 19

Edwards and Lewis v United Kingdom (2005) 40 EHRR 24 26

Factory at Chorzów (Germany v Poland) (Merits) PCIJ Series A No 17 39

Fourie v Le Roux [2007] UKHL 1, [2007] 1 WLR 320 25

Fox, Campbell and Hartley v United Kingdom (1990) 13 EHRR 157 25

Gabčíkovo-Nagymaros Project (Hungary/Slovakia) [1997] ICJ Rep 7 26

Keefe v Isle of Man Steam Packet Co Ltd [2010] EWCA Civ 683 37

Kennedy v Charity Commission [2014] UKSC 20, [2015] AC 455 39

La Quadrature du Net and Others v Premier ministre and Others (Joined Cases C-511/18, C-512/18 and C-520/18) EU:C:2020:791 19

Microsoft Corporation v United States 829 F 3d 197 (2d Cir 2016), vacated as moot 584 US 236 (2018) 20

Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America) (Merits) [1986] ICJ Rep 14 15

Ninemia Maritime Corp v Trave Schiffahrtsgesellschaft mbH (The Niedersachsen) [1983] 1 WLR 1412 (CA) 1422 25

Norwich Pharmacal Co v Customs and Excise Commissioners [1974] AC 133 (HL) 174–75.....25

O'Hara v Chief Constable of the Royal Ulster Constabulary [1997] AC 286 (HL) 298–99.....13

Oil Platforms (Islamic Republic of Iran v United States of America) [2003] ICJ Rep 161.....18

Prosecutor v Tadić (Appeal Judgment) IT-94-1-A (15 July 1999).....30

Pulp Mills on the River Uruguay (Argentina v Uruguay) [2010] ICJ Rep 14.....18

R (Haralambous) v Crown Court at St Albans [2018] UKSC 1, [2018] AC 236.....35

R (KBR Inc) v Director of the Serious Fraud Office [2021] UKSC 2, [2022] AC 519.....20

R v Chief Constable of the West Midlands Police, ex p Wiley [1995] 1 AC 274 (HL) 306–07.....35

R v H [2004] UKHL 3, [2004] 2 AC 134.....17, 26

Re B (Children) (Care Proceedings: Standard of Proof) [2008] UKHL 35, [2009] 1 AC 11.....18

Secretary of State for the Home Department v AF (No 3) [2009] UKHL 28, [2010] 2 AC 269.....35

Shvidler v Secretary of State for Foreign, Commonwealth and Development Affairs [2025] UKSC 30.....24

Tariq v Home Office [2011] UKSC 35, [2012] 1 AC 452.....35

Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Watson (Joined Cases C-203/15 and C-698/15) EU:C:2016:970.....19

United States Diplomatic and Consular Staff in Tehran (United States of America v Iran) (Judgment) [1980] ICJ Rep 3.....15

United States v Yin Kecheng and Others, Indictment, Case No 1:24-cr-00323 (DDC, filed 6 November 2024, unsealed 5 March 2025).....42

Wisniewski v Central Manchester Health Authority [1998] PIQR P324 (CA) P340.....37

Woolmington v Director of Public Prosecutions [1935] AC 462 (HL) 481–82.....13

Youssef v Secretary of State for Foreign and Commonwealth Affairs [2016] UKSC 3, [2016] AC 1457.....24

Legislation

Civil Procedure Rules 1998.....25

Clarifying Lawful Overseas Use of Data Act 2018, Pub L No 115-141, div V, 132 Stat 1213.....25

Computer Misuse Act 1990.....15, 21

Crime (Overseas Production Orders) Act 2019.....19, 25

Criminal Procedure and Investigations Act 1996.....17, 26

Crown Proceedings Act 1947.....39

Cyber Security and Resilience (Network and Information Systems) Bill, HL Bill 32 of 2026–27 (as brought from the Commons, 17 June 2026).....21

Data Protection Act 2018.....21

Economic Crime and Corporate Transparency Act 2023.....15

Freedom of Information Act 2000.....37

Human Rights Act 1998.....39–40

Investigatory Powers Act 2016.....21, 25

Justice and Security Act 2013....35, 37

Network and Information Systems Regulations 2018, SI 2018/506....21

Police and Criminal Evidence Act 1984....13, 25

Regulation of Investigatory Powers Act 2000....35

Russia (Sanctions) (EU Exit) Regulations 2019, SI 2019/855....13, 24

Sanctions and Anti-Money Laundering Act 2018....13, 24

Senior Courts Act 1981....25

Serious Crime Act 2007....15

Special Immigration Appeals Commission Act 1997....35

Special Immigration Appeals Commission (Procedure) Rules 2003, SI 2003/1034....35

Treaties and International Instruments

Agreement between the Government of the United Kingdom of Great Britain and Northern Ireland and the Government of the United States of America on Access to Electronic Data for the Purpose of Countering Serious Crime (signed 3 October 2019, entered into force 3 October 2022) TS No 33/2024....19

Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS XVI....13

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended)....19

Council of Europe Convention on Cybercrime (opened for signature 23 November 2001, entered into force 1 July 2004) ETS No 185....19, 22, 25

International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries' (2001) UN Doc A/56/10....13

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3....28

Second Additional Protocol to the Convention on Cybercrime on Enhanced Co-operation and Disclosure of Electronic Evidence (opened for signature 12 May 2022) CETS No 224....19

Standards, Government and Technical Materials

Attorney General's Office, 'Attorney General's Speech at the International Institute for Strategic Studies' (11 January 2017) https://www.gov.uk/government/speeches/attorney-generals-speech-at-the-international-institute-for-strategic-studies accessed 6 August 2026

Attorney General's Office, 'Cyber and International Law in the 21st Century' (23 May 2018) https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century accessed 6 August 2026

Australian Government, *Australia's Position on How International Law Applies to State Conduct in Cyberspace* (2020)

Cabinet Office, *Professional Standards for Government Intelligence Assessment* (November 2018)

Comptroller and Auditor General, *Investigation: WannaCry Cyber Attack and the NHS* (HC 2017–19, 414) 5–13

Council of the European Union, 'Cyber-attacks: Six Individuals Added to EU Sanctions List for Malicious Cyber Activities against EU Member States and Ukraine' (24 June 2024)

Council of the European Union, 'Russian Cyber Operations against Ukraine: Declaration by the High Representative on behalf of the European Union' (10 May 2022) https://www.consilium.europa.eu/en/press/press-releases/2022/05/10/russian-cyber-operations-against-ukraine-declaration-by-the-high-representative-on-behalf-of-the-european-union/ accessed 6 August 2026

Crown Prosecution Service, ‘Cybercrime—Prosecution Guidance’ https://www.cps.gov.uk/prosecution-guidance/cybercrime-prosecution-guidance accessed 6 August 2026

Crown Prosecution Service, *Director’s Guidance on Charging* (6th edn, December 2020, revised March 2025)

Crown Prosecution Service, *The Code for Crown Prosecutors* (26 October 2018)

Cybercrime Convention Committee, *24/7 Points of Contact Network of the Budapest Convention on Cybercrime: Assessment Report* (T-CY(2020)2, 2020)

Cybersecurity and Infrastructure Security Agency, ‘Advanced Persistent Threat Compromise of Government Agencies, Critical Infrastructure, and Private Sector Organizations’ (AA20-352A, updated 15 April 2021) https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-352a accessed 6 August 2026

Cybersecurity and Infrastructure Security Agency and others, *Best Practices for Event Logging and Threat Detection* (21 August 2024) https://www.cisa.gov/resources-tools/resources/best-practices-event-logging-and-threat-detection accessed 6 August 2026

Cybersecurity and Infrastructure Security Agency and others, ‘Enhanced Visibility and Hardening Guidance for Communications Infrastructure’ (3 December 2024) https://www.cisa.gov/resources-tools/resources/enhanced-visibility-and-hardening-guidance-communications-infrastructure accessed 6 August 2026

Cybersecurity and Infrastructure Security Agency and others, *Primary Mitigations to Reduce Cyber Threats to Operational Technology* (6 May 2025) https://www.cisa.gov/resources-tools/resources/primary-mitigations-reduce-cyber-threats-operational-technology accessed 6 August 2026

Cybersecurity and Infrastructure Security Agency, *MAR-17-352-01 HatMan—Safety System Targeted Malware* (12 April 2018)

Department for Energy Security and Net Zero, *National Emergency Plan 2023: Downstream Gas and Electricity* (July 2023) https://www.gov.uk/government/publications/national-emergency-plan-downstream-gas-and-electricity-2016 accessed 6 August 2026

Department for Science, Innovation and Technology, *AI Cyber Security Code of Practice* (31 January 2025)

Department for Science, Innovation and Technology, ‘Summary of the Bill’ (30 June 2026) https://www.gov.uk/government/publications/cyber-security-and-resilience-network-and-information-systems-bill-factsheets/summary-of-the-bill accessed 6 August 2026

Department of Health and Social Care, ‘Synnovis Cyber Attack’ (Written Statement HCWS1046, 30 June 2025)

Federal Government of Germany, ‘On the Application of International Law in Cyberspace’ (March 2021)

Foreign, Commonwealth & Development Office, *Post-Legislative Scrutiny Memorandum: Sanctions and Anti-Money Laundering Act 2018* (CP 1020, March 2024) paras 102–19.....24

Foreign, Commonwealth & Development Office, ‘Application of International Law to States’ Conduct in Cyberspace: UK Statement’ (3 June 2021) https://www.gov.uk/government/publications/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement accessed 6 August 2026

Foreign, Commonwealth & Development Office, *The Pall Mall Process Code of Practice for States* (4 April 2025) https://www.gov.uk/government/publications/the-pall-mall-process-code-of-practice-for-states accessed 6 August 2026

Foreign, Commonwealth & Development Office, ‘UK Holds China State-affiliated Organisations and Individuals Responsible for Malicious Cyber Activity’ (25 March 2024) https://www.gov.uk/government/news/uk-holds-china-state-affiliated-organisations-and-individuals-responsible-for-malicious-cyber-activity accessed 6 August 2026

Foreign & Commonwealth Office and National Cyber Security Centre, ‘Foreign Office Minister Condemns Russia for NotPetya Attacks’ (15 February 2018) https://www.gov.uk/government/news/foreign-office-minister-condemns-russia-for-notpetya-attacks accessed 6 August 2026

France, Ministry of the Armies, *International Law Applied to Operations in Cyberspace* (2019)

Government of Canada, *International Law Applicable in Cyberspace* (22 April 2022)

Government of the Netherlands, 'Letter to the Parliament on the International Legal Order in Cyberspace' (5 July 2019)

HM Treasury, *OFSI Guidance for the UK Financial Sanctions Regime* (December 2025) 20–35

Home Office, *Equipment Interference Code of Practice* (December 2022)

ICRC, *International Humanitarian Law and Cyber Operations during Armed Conflicts* (Position Paper, November 2019) 4–8

ICRC, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (Nils Melzer ed, ICRC 2009) 65–73

Intelligence Analysis, 'Explaining Uncertainty in UK Intelligence Assessment' (24 March 2025) https://www.gov.uk/government/publications/explaining-uncertainty-in-uk-intelligence-assessment/explaining-uncertainty-in-uk-intelligence-assessment accessed 6 August 2026

Intelligence Analysis, 'PHIA Common Analytical Standards' (24 March 2025) https://www.gov.uk/government/publications/phia-common-analytical-standards/phia-common-analytical-standards accessed 6 August 2026

Intelligence and Security Committee of Parliament, *Russia* (HC 632, 2019–21)

International Organization for Standardization, *ISO/IEC 27037:2012—Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence* (2012)

Iraq Inquiry, *The Report of the Iraq Inquiry: Executive Summary* (HC 264, 2016–17)

Judicial College, *Crown Court Compendium Part I: Jury and Trial Management and Summing Up* (October 2025) pt 5-1 https://www.judiciary.uk/wp-content/uploads/2022/06/Crown-Court-Compendium-Part-I.pdf accessed 6 August 2026

Liu F and others, *NIST Cloud Computing Reference Architecture* (NIST SP 500-292, September 2011) 5–13

Ministry of Defence, *Intelligence, Counter-intelligence and Security Support to Joint Operations* (Joint Doctrine Publication 2-00, 4th edn, October 2023)

Ministry of Defence, *The Manual of the Law of Armed Conflict* (OUP 2004)

MITRE, 'Groups' *MITRE ATT&CK* https://attack.mitre.org/groups/ accessed 6 August 2026

National Cyber Security Centre and others, *Guidelines for Secure AI System Development* (November 2023)

National Cyber Security Centre and others, 'PRC State-Sponsored Actors Compromise and Maintain Persistent Access to US Critical Infrastructure' (7 February 2024) https://www.cisa.gov/news-events/cybersecurity-advisories/aa24-038a accessed 6 August 2026

National Cyber Security Centre, *Cyber Incident Response* https://www.ncsc.gov.uk/section/about-ncsc/incident-management accessed 6 August 2026

National Cyber Security Centre, 'Logging Made Easy' (updated 18 September 2025) https://www.ncsc.gov.uk/information/logging-made-easy accessed 6 August 2026

National Cyber Security Centre, *The Near-term Impact of AI on the Cyber Threat* (24 January 2024) 4–9 https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat accessed 6 August 2026

National Cyber Security Centre, 'UK Calls Out China State-affiliated Actors for Malicious Cyber Targeting of UK Democratic Institutions and Parliamentarians' (25 March 2024) https://www.ncsc.gov.uk/news/china-state-affiliated-actors-target-uk-democratic-institutions-parliamentarians accessed 6 August 2026

National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1, January 2023) 17–31

National Institute of Standards and Technology, *Guide to Computer Security Log Management* (NIST SP 800-92, September 2006) 2-1–2-6

National Institute of Standards and Technology, *Guide to Integrating Forensic Techniques into Incident Response* (NIST SP 800-86, August 2006) 3-1–3-11

National Institute of Standards and Technology, *Incident Response Recommendations and Considerations for Cybersecurity Risk Management* (NIST SP 800-61 rev 3, April 2025)

National Institute of Standards and Technology, *NIST Cloud Computing Forensic Reference Architecture* (NIST SP 800-201, July 2024) 1–9

National Institute of Standards and Technology, *NIST Cloud Computing Forensic Science Challenges* (NISTIR 8006, June 2014) 12–38

National Institute of Standards and Technology, *Secure Software Development Framework* (NIST SP 800-218, February 2022)

National Telecommunications and Information Administration, *The Minimum Elements for a Software Bill of Materials* (12 July 2021) 5–15

New Zealand Ministry of Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace* (1 December 2020)

NHS England, 'Synnovis Cyber Incident' https://www.england.nhs.uk/synnovis-cyber-incident/ accessed 6 August 2026

Office of the Director of National Intelligence, *Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution* (6 January 2017)

Office of the Director of National Intelligence, *Intelligence Community Directive 203: Analytic Standards* (21 December 2022) 3–6 https://www.dni.gov/files/documents/ICD/ICD-203.pdf accessed 6 August 2026

Office of the Director of National Intelligence, *Intelligence Community Directive 208: Maximizing the Utility of Analytic Products* (9 January 2017) 1–3 https://www.dni.gov/files/documents/ICD/ICD-208-Maximizing-the-Utility-of-Analytic-Products-2017-01-09.pdf accessed 6 August 2026

Office of the Director of National Intelligence, *Intelligence Community Directive 505: Artificial Intelligence* (20 January 2025) 8–10 https://www.dni.gov/files/documents/ICD/ICD-505-Artificial-Intelligence.pdf accessed 6 August 2026

Organization for Security and Co-operation in Europe, Decision No 1202, 'OSCE Confidence-Building Measures to Reduce the Risks of Conflict Stemming from the Use of Information and Communication Technologies' (10 March 2016) PC.DEC/1202

Review of Intelligence on Weapons of Mass Destruction, *Report of a Committee of Privy Counsellors* (HC 898, 2003–04)

Sherman Kent, 'Words of Estimative Probability' (1964) 8(4) *Studies in Intelligence* 49, 49–65 https://www.cia.gov/resources/csi/studies-in-intelligence/archives/vol-8-no-4/words-of-estimative-probability/ accessed 6 August 2026

SLSA, 'Supply-chain Levels for Software Artifacts Specification v1.0' (2023) https://slsa.dev/spec/v1.0/ accessed 6 August 2026

Swiss Federal Department of Foreign Affairs and ICRC, *Montreux Document on Pertinent International Legal Obligations and Good Practices for States related to Operations of Private Military and Security Companies during Armed Conflict* (17 September 2008)

UNGA, 'Official Compendium of Voluntary National Contributions on the Subject of How International Law Applies to the Use of Information and Communications Technologies by States' (13 July 2021) UN Doc A/76/136

UNGA, 'Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security: Final Substantive Report' (10 March 2021) UN Doc A/AC.290/2021/CRP.2,

US Department of Health and Human Services, Office for Civil Rights, 'Cyberattack on Change Healthcare' (Dear Colleague Letter, 13 March 2024) https://www.hhs.gov/sites/default/files/cyberattack-change-healthcare.pdf accessed 6 August 2026

US Department of Justice, 'Justice Department Charges 12 Chinese Contract Hackers and Law Enforcement Officers in Global Computer Intrusion Campaigns' (5 March 2025) https://www.justice.gov/opa/pr/justice-department-charges-12-chinese-contract-hackers-and-law-enforcement-officers-global accessed 6 August 2026

US Department of Justice, 'Seven Hackers Associated with Chinese Government Charged with Computer Intrusions Targeting Perceived Critics of China and US Businesses and Political Officials' (25 March 2024) https://www.justice.gov/opa/pr/seven-hackers-associated-chinese-government-charged-computer-intrusions-targeting accessed 6 August 2026

US Department of Justice, ‘Six Russian GRU Officers Charged in Connection with Worldwide Deployment of Destructive Malware and Other Disruptive Actions in Cyberspace’ (19 October 2020) https://www.justice.gov/archives/opa/pr/six-russian-gru-officers-charged-connection-worldwide-deployment-destructive-malware-and accessed 6 August 2026

US Department of Justice, ‘US Government Disrupts Botnet People’s Republic of China Used to Conceal Hacking of Critical Infrastructure’ (31 January 2024) https://www.justice.gov/opa/pr/us-government-disrupts-botnet-peoples-republic-china-used-conceal-hacking-critical accessed 6 August 2026

US Senate Committee on Finance, *Hacking America’s Health Care: Assessing the Change Healthcare Cyber Attack and What’s Next* (1 May 2024)

Corporate, Technical and Contemporary Materials

Google Threat Intelligence Group, *Adversarial Misuse of Generative AI* (January 2025) 4–28 https://cloud.google.com/resources/content/adversarial-misuse-of-generative-ai accessed 6 August 2026

UnitedHealth Group, *Change Healthcare Cyber Response Update* (7 May 2024) https://www.unitedhealthgroup.com/ns/changehealthcare.html accessed 6 August 2026

1. The Missing Law of Decision

The previous papers identify a species of hostile power which existing categories divide too soon. Private capital can supply strategic communications, compute, targeting and intrusion; a PMSC can accept an objective whilst subcontractors and an autonomous agent select the means; a controller can shard execution across jurisdictions so that no territorial State sees the campaign; MOSAIC can correlate the resulting fragments; and a statute can confer powers of preservation, restraint, sanction and immediate defence. None of that answers the question which confronts the minister at two in the morning: how much of the assembled account must be believed, and in what form, before this particular power may be used?[1]

The common answer—‘high confidence’—is inadequate. It may describe an analyst’s confidence in the quality of the source base, a probabilistic judgement that a proposition is true, an institutional consensus, or merely the author’s rhetorical conviction. It does not identify the proposition. One team may be highly confident that two intrusions share an operator; another that the infrastructure was procured by a contractor; a third that a State service paid the contractor; and a lawyer may nevertheless conclude that the evidence does not establish instruction or control in the operation complained of. Multiplying those judgements into one percentage produces an appearance of precision at the cost of the very distinctions upon which responsibility depends.[2]

Intelligence practice itself warns against this collapse. The United Kingdom’s professional standards require information to be separated from assumptions, uncertainty to be expressed, source quality and denial or deception to be considered, disagreements to be recorded and significant changes from earlier assessments to be explained. The United States’ analytic standards similarly distinguish likelihood from confidence in the source base and require alternative explanations to be addressed. Sherman Kent’s

[1] Teoman M Hagemeyer-Witzleb, *The International Law of Economic Warfare* (Springer 2021) 1–17; Tim Maurer, *Cyber Mercenaries: The State, Hackers, and Power* (CUP 2018) 1–18; Evelyne Schmid and Ayşe Özge Erceiş, ‘The Attribution of Omissions: Due Diligence in Cyberspace and State Responsibility’ (2023) 33 *Swiss Review of International and European Law* 577, 581–90; National Institute of Standards and Technology, *NIST Cloud Computing Forensic Reference Architecture* (NIST SP 800-201, July 2024) 1–9.

[2] Thomas Rid and Ben Buchanan, ‘Attributing Cyber Attacks’ (2015) 38 *Journal of Strategic Studies* 4, 14–23 https://doi.org/10.1080/01402390.2014.977382; Sherman Kent, ‘Words of Estimative Probability’ (1964) 8(4) *Studies in Intelligence* 49, 49–65 https://www.cia.gov/resources/csi/studies-in-intelligence/archives/vol-8-no-4/words-of-estimative-probability/ accessed 6 August 2026; Office of the Director of National Intelligence, *Intelligence Community Directive 203: Analytic Standards* (21 December 2022) 3–6 https://www.dni.gov/files/documents/ICD/ICD-203.pdf accessed 6 August 2026; Intelligence Analysis, ‘Explaining Uncertainty in UK Intelligence Assessment’ (24 March 2025) https://www.gov.uk/government/publications/explaining-uncertainty-in-uk-intelligence-assessment/explaining-uncertainty-in-uk-intelligence-assessment accessed 6 August 2026.

older complaint remains exact: verbal probabilities fail where writer and reader attach different odds to the same word.[3] The cure is not to abolish judgement, for attribution cannot be reduced to a checksum; it is to make the judgement auditable.

Cyber attribution is neither uniquely impossible nor ordinarily instantaneous. Rid and Buchanan describe it as an all-source analytical process rather than a merely technical trace. Eichensehr demonstrates that public attribution is at once technical, legal and political, frequently involving companies and allied governments as well as intelligence services. Egloff's work upon public attribution further shows that the act is relational: an accuser chooses what to reveal, which partners to assemble and what consequence to attach.[4] The difficulty is therefore not an excuse for inaction. It is a reason to specify the decision.

The law already employs different evidential gates according to function. A constable may arrest upon reasonable grounds for suspicion; a prosecutor requires a realistic prospect of conviction; a criminal court must be sure. A minister designating a person under sanctions legislation applies the statutory reasonable-grounds test and remains subject to legality and proportionality. A civil court considering interim relief asks whether there is a serious issue, what course carries the lower risk of injustice and where the balance lies. International responsibility, countermeasures and self-defence ask different questions again.[5] There is no embarrassment in plurality. The embarrassment lies in speaking as though one undifferentiated attribution settled them all.

Nor does gravity always demand delay. The prospect of catastrophic harm may justify immediate evidence preservation, isolation of a credential, warning to a victim or suspension of one malicious workload upon material far below that required for punishment. Reversibility, scope, duration and the consequence of error matter. An order which retains volatile logs for twenty-four hours is not the same act as a public accusation; freezing one suspicious transaction is not confiscation; blocking a command channel is not killing the person believed to control it. The legal system is capable of acting upon provisional fact because it pairs provisional authority with a clock, a narrow object and review.

Yet the same danger must not be permitted to reason in a circle. 'The consequences could be terrible; therefore State X probably did it; therefore terrible consequences are likely' is not analysis. Maximum reasonably credible foreseeable harm belongs to urgency, precaution, defensive scope and necessity. It cannot transform a weak identity inference into proof of authorship, still less prove State control. The greater the irreversibility of the proposed response, the more carefully the decision maker must separate confidence in the harm proposition from confidence in the actor, legal relationship and target propositions.

---

[3] Intelligence Analysis, 'PHIA Common Analytical Standards' (24 March 2025) https://www.gov.uk/government/publications/phia-common-analytical-standards/phia-common-analytical-standards accessed 6 August 2026; Office of the Director of National Intelligence (n 2) 3–6; Kent (n 2) 49–65; Ministry of Defence, *Intelligence, Counter-intelligence and Security Support to Joint Operations* (Joint Doctrine Publication 2-00, 4th edn, October 2023) paras 3.44–3.47.

[4] Rid and Buchanan (n 2) 4–37; Kristen E Eichensehr, 'The Law and Politics of Cyberattack Attribution' (2020) 67 *UCLA Law Review* 520, 528–46; Florian J Egloff and Max Smeets, 'Publicly Attributing Cyber Attacks: A Framework' (2023) 46 *Journal of Strategic Studies* 502, 502–33 https://doi.org/10.1080/01402390.2021.1895117; Florian J Egloff, 'Public Attribution of Cyber Intrusions' (2020) 6 *Journal of Cybersecurity* tyaa012 https://doi.org/10.1093/cybsec/tyaa012; Nicholas Tsagourias and Fiona Middleton, 'Fact-Finding and Cyber Attribution' in Russell Buchan, Daniel Franchini and Nicholas Tsagourias (eds), *The Changing Character of International Dispute Settlement: Challenges and Prospects* (CUP 2023) 439–66 https://doi.org/10.1017/9781009076296.023.

[5] Police and Criminal Evidence Act 1984, s 24; *O'Hara v Chief Constable of the Royal Ulster Constabulary* [1997] AC 286 (HL) 298–99; Crown Prosecution Service, *The Code for Crown Prosecutors* (26 October 2018) paras 4.6–4.8 https://www.cps.gov.uk/publication/code-crown-prosecutors accessed 6 August 2026; *Woolmington v Director of Public Prosecutions* [1935] AC 462 (HL) 481–82; Sanctions and Anti-Money Laundering Act 2018, ss 11–12 and 23; Russia (Sanctions) (EU Exit) Regulations 2019, SI 2019/855, reg 6; *Bank Mellat v HM Treasury (No 2)* [2013] UKSC 39, [2014] AC 700, [68]–[76]; *American Cyanamid Co v Ethicon Ltd* [1975] AC 396 (HL) 406–09; International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries' (2001) UN Doc A/56/10, arts 49–54; Charter of the United Nations (adopted 26 June 1945, entered into force 24 October 1945) 1 UNTS XVI, art 51.

This paper accordingly rejects two extremes. The first demands courtroom proof before government may preserve evidence or prevent a machine from completing an attack. That rule would make effective defence unlawful until it was unnecessary. The second treats national-security belief as a universal warrant. That rule mistakes executive responsibility for evidential immunity. The proper model is staged: each response has its own factual proposition, legal gateway, threshold, record and route of correction.

The model is especially necessary where artificial intelligence or private intermediaries disaggregate the act. A State may pay a prime contractor for an outcome; engineers in several time zones may write components; an integrator may connect an objective-bearing agent to rented access, stolen identity and cloud capacity; the agent may choose technique, target and route. Technical attribution may reach the workload. Criminal evidence may identify the integrator. Contract and payment may identify the payer. International law may or may not attribute the agent's conduct to a State. A lawful response cannot be selected until the decision maker states which of those propositions the measure actually requires.

The distinction between the technical and the legal questions is not new and this paper does not claim it. Tsagourias and Farrell make that division the organising frame of their account of attribution, canvassing the three families of proposal which have followed from it: an institutional attribution mechanism, revision of the legal determinants of attribution, and a standard of proof. Roscini had earlier set out the evidentiary problem in disputes over State responsibility for cyber operations, and Finnemore and Hollis have since examined public accusation as a practice which makes law rather than merely reporting it. What none of that work supplies, because none of it was attempting to, is the third question. Technical and legal attribution ask what happened and whose act it was in law. Neither asks what a decision maker must be satisfied of before exercising a particular power, which is a question with a different answer for a preservation notice, a sanctions designation, a criminal charge, a countermeasure and a use of force. The contribution offered here is that third layer and the ladder which follows from it, and the claim is confined to it.[6]

2. Attribution Is Three Questions, Not One

2.1 Technical Attribution

Technical attribution asks what happened and how the operation was constituted. Its propositions include initial access, exploited vulnerability, credential use, command infrastructure, malware lineage, compilation and deployment history, lateral movement, target selection, exfiltration, effect, persistence and attempted concealment. At the most granular level it may identify a device or process. At a wider level it may connect events into an intrusion set or campaign. Neither level, without more, names the legally responsible person.

The evidence is heterogeneous. Logs record an event from the perspective and clock of one system. Packet captures record traffic, not necessarily the human or process ultimately directing it. Malware may contain code similarity, language artefacts, working hours and reused certificates; each can be planted, inherited or shared. Registration, payment and hosting records may identify an account whose credentials were stolen. Victim telemetry is selected by collection rules and retention periods. Human intelligence may explain purpose whilst remaining unavailable for public disclosure. The operation may

[6] Rid and Buchanan (n 2) 14–23; National Cyber Security Centre and others, 'PRC State-Sponsored Actors Compromise and Maintain Persistent Access to US Critical Infrastructure' (7 February 2024) https://www.cisa.gov/news-events/cybersecurity-advisories/aa24-038a accessed 6 August 2026; National Institute of Standards and Technology, *NIST Cloud Computing Forensic Science Challenges* (NISTIR 8006, June 2014) 12–38; Liu F and others, *NIST Cloud Computing Reference Architecture* (NIST SP 500-292, September 2011) 5–13.

also use compromised routers and small-office devices precisely to ensure that the visible source belongs to an innocent third party.[7]

Forensic standards therefore emphasise identification, collection, acquisition, preservation and an auditable chain of custody. Incident-handling guidance separates preparation, detection, containment, eradication and recovery, whilst log-management standards expose how quickly the evidential record can disappear. These are not clerical concerns. If the clock was wrong, the image altered, the volatile memory lost or the collection rule unknown, an apparently exact sequence may be incapable of supporting the inference placed upon it.[8]

The technical conclusion should consequently be expressed as a set of propositions rather than an actor label. It should say, for example: the same controller probably operated infrastructure A and B; tool C was deployed from B; account D authorised the target query; D was funded through payer E; and two independent sources connect E to organisation F. Each proposition then carries its own confidence, provenance, alternative and correction history. A later discovery which disproves the payment link need not erase the sound conclusion concerning common control of A and B.

This proposition model resists the familiar ‘APT’ fallacy. A threat-cluster name is an analytical convenience whose boundary depends upon the vendor’s visibility and method. Two companies may divide or combine activity differently; a government may adopt another label; an adversary may borrow tools; a contractor may work for several customers. To say that an intrusion resembles a named cluster is useful. To convert that resemblance silently into proof of one State’s direction is not.[9]

2.2 Legal Attribution

Legal attribution asks a different question: whose conduct is it in law? For individual criminal responsibility the inquiry concerns actus reus, mens rea, participation, assistance, conspiracy and the territorial or nationality bases of jurisdiction. For a company it may concern directing mind, identification, vicarious or statutory corporate responsibility and failure-to-prevent rules. For international responsibility it concerns, amongst other things, a State organ, an entity empowered to exercise governmental authority, instructions, direction or control, adoption, acknowledgment or knowing assistance.[10]

The difference is decisive. Proof that an operator is a citizen of State X does not make the operation an act of X. Proof that infrastructure stood in X does not establish that X knew of it. Proof that an

---

[7] International Organization for Standardization, *ISO/IEC 27037:2012—Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence* (2012); National Institute of Standards and Technology, *Guide to Integrating Forensic Techniques into Incident Response* (NIST SP 800-86, August 2006) 3-1–3-11; National Institute of Standards and Technology, *Guide to Computer Security Log Management* (NIST SP 800-92, September 2006) 2-1–2-6; National Institute of Standards and Technology, *Incident Response Recommendations and Considerations for Cybersecurity Risk Management* (NIST SP 800-61 rev 3, April 2025); Cybersecurity and Infrastructure Security Agency and others, *Best Practices for Event Logging and Threat Detection* (21 August 2024) https://www.cisa.gov/resources-tools/resources/best-practices-event-logging-and-threat-detection accessed 6 August 2026; National Cyber Security Centre, ‘Logging Made Easy’ (updated 18 September 2025) https://www.ncsc.gov.uk/information/logging-made-easy accessed 6 August 2026.

[8] Office of the Director of National Intelligence, *Background to “Assessing Russian Activities and Intentions in Recent US Elections”: The Analytic Process and Cyber Incident Attribution* (6 January 2017); MITRE, ‘Groups’ *MITRE ATT&CK* https://attack.mitre.org/groups/ accessed 6 August 2026; National Cyber Security Centre, *The Near-term Impact of AI on the Cyber Threat* (24 January 2024) 4–9 https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat accessed 6 August 2026; Rid and Buchanan (n 2) 14–23.

[9] Computer Misuse Act 1990, ss 1–3ZA; Serious Crime Act 2007, ss 44–46; Economic Crime and Corporate Transparency Act 2023, ss 196 and 199–206; International Law Commission (n 5) arts 4–5, 8, 11 and 16; James Crawford, *State Responsibility: The General Part* (CUP 2013) 113–66; Helmut Philipp Aust, *Complicity and the Law of State Responsibility* (CUP 2011) 191–249.

[10] International Law Commission (n 5) arts 4, 7 and 8; *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America)* (Merits) [1986] ICJ Rep 14, [109]–[115]; *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro)* [2007] ICJ Rep 43, [385]–[407]; Crawford (n 10) 115–31.

intelligence officer paid a contractor may be powerful evidence, but the applicable attribution rule still asks what was instructed, directed or controlled in relation to the impugned conduct. Conversely, a State organ does not cease to be an organ when it uses a personal device, false corporate identity or foreign server. Legal attribution follows status and relationship; it is not defeated or established by the last IP address.[11]

International law also distinguishes attribution of conduct from breach, knowledge and remedy. A State may be responsible for its organ's operation although the operation violates no primary rule. It may breach a due-diligence obligation through its own omission without the attackers' conduct becoming its act. It may assist another State's wrongful act under article 16 without becoming the principal author. It may later acknowledge and adopt private conduct. Those routes have different factual predicates. A single verdict of 'State-sponsored' tells the decision maker almost nothing about which is proved.[12]

The distinction is equally important in public speech. 'Associated with', 'linked to', 'State-aligned', 'State-sponsored', 'working on behalf of' and 'attributable to' are not a ladder whose top rung is always article 8 control. Official statements often use policy language deliberately broader or narrower than a pleaded legal conclusion. A publication may properly identify a technical and organisational relationship without asserting that every act in a campaign is attributable to the State under the Articles on Responsibility of States. The decision record must not improve the source by paraphrase.

2.3 Decisional Attribution

Decisional attribution asks whether the proposition required for the proposed response is sufficiently established. It is the bridge between fact and power. The proposition changes with the measure. Preserving a provider's logs may require reasonable grounds to believe that the records are material and at risk. Isolating a workload may require reasonable grounds to believe that it is presently enabling hostile conduct. Sanctioning a payer requires the statutory connection to the sanctioned activity. Prosecuting an engineer requires admissible evidence of the offence and mental element. Taking a countermeasure against a State requires a prior breach attributable to that State. Using force against a person or object requires the facts necessary to establish self-defence and lawful targeting.

The same evidence may satisfy one and fail another without contradiction. A classified intercept may give government compelling intelligence that a State service commissioned an operation, yet be unusable at a criminal trial because disclosure would reveal a source and no substitute proof exists. Public technical reporting may justify a warning and coordinated hardening whilst remaining insufficient to freeze a named person's property. A payment record may prove a contractor relationship but not the customer's instruction to attack this hospital. Decisional attribution makes those limits visible.

It follows that the threshold must be attached to a proposition and a consequence. 'High confidence in Russia' is not a legal standard. 'The decision maker is highly confident, upon two independent sources and authenticated payment records, that entity P knowingly financed the continuing operation, and regulation R permits temporary designation upon reasonable grounds to suspect that connection' is capable of audit. Whether the decision is right remains contestable; what was decided is at least intelligible.

Decisional attribution has a temporal dimension. The lawful question at 02.00 may concern preservation; at 02.15 containment; at 08.00 public warning; after a week sanctions; after disclosure

[11] International Law Commission (n 5) arts 2, 4, 8, 11 and 16 and commentaries; Schmid and Erceiş (n 1) 581–90; Aust (n 10) 191–249; *United States Diplomatic and Consular Staff in Tehran (United States of America v Iran)* (Judgment) [1980] ICJ Rep 3, [58]–[74].

[12] ISO/IEC 27037:2012 (n 8); Criminal Procedure and Investigations Act 1996, ss 3 and 7A; *R v H* [2004] UKHL 3, [2004] 2 AC 134, [13]–[36]; PHIA Common Analytical Standards (n 3); Office of the Director of National Intelligence (n 2) 4–6.

review a charge. Evidence and risk change. A system which overwrites the earlier conclusion with the latest one destroys accountability in both directions. It prevents an official from showing why urgent action was reasonable upon the material then available, and prevents the affected person from showing that government persisted after the basis had collapsed.

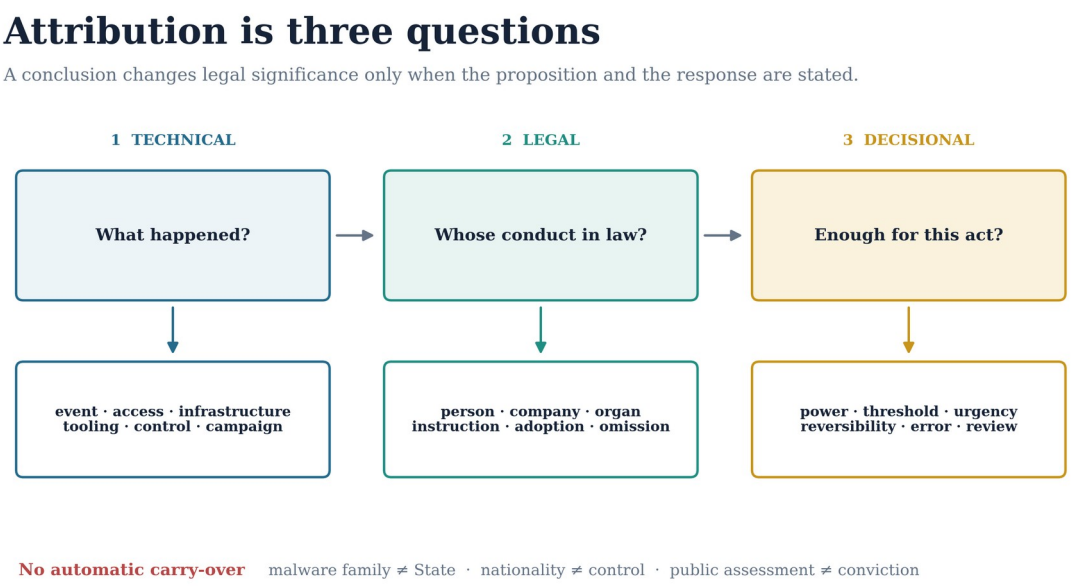

Figure 1. Attribution consists of technical, legal and decisional propositions. No conclusion in one column automatically supplies the next.

3. The Evidential Object

3.1 A Claim Graph, Not a Dossier Label

The basic unit of attribution should be a claim whose support can be traced. A claim graph contains nodes—event, infrastructure, account, person, organisation, State organ, payer, target and effect—and edges which state the asserted relationship. An edge may be direct observation, common control, payment, employment, instruction, technical dependence, legal status or temporal sequence. The graph does not decide the law. It prevents an analyst from moving between meanings of ‘link’ without recording the move.

Five attributes should accompany each material edge. First is provenance: who collected the material, by what authority and with what chain of custody? Secondly is independence: does a second report derive from a separate source, or merely repeat the first? Thirdly is reliability: was the system complete, the witness placed to know, the method validated and the record protected? Fourthly is deception opportunity: who could plant, alter, replay or route the indicator? Fifthly is legal usability: may the material be shown to the decision maker, disclosed to a court, shared with an ally and relied upon for the contemplated purpose?[13]

[13] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1, January 2023) 17–31; National Cyber Security Centre and others, *Guidelines for Secure AI System Development* (November 2023); Department for Science, Innovation and Technology, *AI Cyber Security Code of Practice* (31 January 2025); National Institute of Standards and Technology, *Secure Software Development Framework* (NIST SP 800-218, February 2022); Office of the Director of National Intelligence, *Intelligence Community Directive 505: Artificial Intelligence* (20 January 2025) 8–10 https://www.dni.gov/files/documents/ICD/ICD-505-Artificial-Intelligence.pdf accessed 6 August 2026.

The distinction between independence and multiplicity is vital. Ten reports derived from one vendor's telemetry are one evidential root. Three allied statements based upon the same undisclosed intercept do not corroborate one another merely because their letterheads differ. A malware hash repeated across agencies proves common observation of the hash, not independent proof of authorship. The claim graph should record source ancestry so that agreement does not become artificial weight.

The same discipline applies to machine analysis. A model may cluster infrastructure, translate code, detect stylistic similarity or rank hypotheses; it remains an analytical instrument. The version, prompt or task, training or retrieval boundary, input set and relevant validation should be preserved where output materially influences the conclusion. A model which confidently completes a familiar pattern may be most dangerous precisely when an adversary has constructed that pattern as a false flag. Government standards for AI risk and secure system development reinforce the need for documented data, testing, human responsibility and monitoring.[14]

3.2 Direct Fact, Inference and Legal Conclusion

Every attribution assessment should separate five species of statement:

1. **direct fact**—the sensor recorded a connection; the provider authenticated an account; the bank recorded a payment;
2. **technical inference**—the infrastructure was probably under common control; the operator reused a tool; the command arrived through a particular route;
3. **organisational inference**—the operator acted for a contractor, unit or customer;
4. **legal conclusion**—the conduct is attributable to a State, satisfies an offence or falls within a designation criterion; and
5. **policy judgement**—public attribution, restraint or response is expedient and lawful now.

This separation does not belittle inference. Corfu Channel remains authority that exclusive territorial control may justify a more liberal recourse to inference where direct proof is unavailable, provided that the indirect evidence leaves no reasonable room for another conclusion upon the point being proved. International courts routinely evaluate circumstantial evidence; domestic courts do the same. The discipline lies in identifying the inferential step and its alternative, not in pretending that only eyewitness testimony is evidence.[15]

Neither should gravity be confused with a free-standing standard of proof. English civil law applies the balance of probabilities even to serious allegations, whilst recognising that inherent probability affects the evidence needed to persuade. Criminal conviction demands that the tribunal be sure. The International Court of Justice has required 'fully conclusive' evidence for charges of exceptional gravity such as genocide, yet has used different formulations in other disputes. These standards arise within distinct jurisdictions and proceedings. They cannot be mixed into a home-made percentage and applied indiscriminately to executive cyber decisions.[16]

The paper therefore uses verbal standards only when anchored to their legal source. 'Reasonable grounds' requires objective, particularised material capable of supporting the specified suspicion or belief. 'Balance of probabilities' asks whether the proposition is more likely than not. 'Realistic

[14] *Corfu Channel (United Kingdom v Albania)* (Merits) [1949] ICJ Rep 4, 18 and 22; *Bosnian Genocide* (n 11) [204]–[210]; Tsagourias and Middleton (n 4) 447–57; *Oil Platforms (Islamic Republic of Iran v United States of America)* [2003] ICJ Rep 161, [57]–[72].

[15] *Re B (Children) (Care Proceedings: Standard of Proof)* [2008] UKHL 35, [2009] 1 AC 11, [2], [13] and [70]–[72]; *Woolmington* (n 5) 481–82; *Bosnian Genocide* (n 11) [209] and [401]–[407]; *Pulp Mills on the River Uruguay (Argentina v Uruguay)* [2010] ICJ Rep 14, [162]–[164].

[16] Rid and Buchanan (n 2) 14–23 and 31–34; Ben Buchanan, *The Cybersecurity Dilemma: Hacking, Trust and Fear between Nations* (OUP 2017) 31–64; Egloff, 'Public Attribution of Cyber Intrusions' (n 4).

prospect of conviction' asks whether an objective and properly directed tribunal is more likely than not to convict upon admissible, reliable and credible evidence. 'Sure' is the criminal trial standard. 'High degree of confidence' in this paper is not a new burden of proof; it is a requirement of decision discipline where international law supplies no verbal formula but the consequence—especially force—is grave and error may be irreversible.

3.3 Contrary Material and Live Hypotheses

Attribution becomes advocacy when it collects only confirmatory indicators. The assessment should state the leading hypothesis, the strongest material against it and at least one reasonably available alternative. It should record whether the alternative was tested and what observation would distinguish the two. This is not a demand to indulge every fantasy. A hypothesis unsupported by known capability, access or motive need not receive equal weight; it should not be dismissed merely because the consequence would be inconvenient.

Deception must be treated as a capability, not an incantation. The question is not whether a false flag is imaginable, for it always is; it is whether the adversary possessed and used a feasible means to produce the observed indicators, and whether any indicator is difficult for an outsider to replicate. Access to a private repository, long-term command history, non-public victim knowledge, authenticated tasking or payment and human reporting may carry weight different in kind from a keyboard layout or compiler timestamp. Rid and Buchanan's model is useful precisely because it treats attribution as accumulating political, operational and technical knowledge rather than hunting for an unforgeable artefact.[17]

A dissent is evidence about the assessment process. The record should preserve which agency or analyst dissented, upon what proposition and because of what source or method, so far as security permits. Consensus reached by deleting the scope of disagreement is weaker than a majority conclusion which identifies it. The Butler Review and Iraq Inquiry demonstrate the cost of intelligence language which becomes firmer as it passes towards policy, while caveat and uncertainty recede.[18]

3.4 Preservation, Correction and the Evidential Clock

Digital evidence decays by design. Cloud instances terminate, volatile memory clears, leases rotate, providers apply short retention periods and autonomous workloads migrate. Preservation is consequently part of attribution, not an administrative afterthought. The first lawful act may properly secure a log, snapshot a workload, retain a payment record, preserve a key and record the current analytical state before any public allegation is possible.[19]

---

[17] Review of Intelligence on Weapons of Mass Destruction, *Report of a Committee of Privy Counsellors* (HC 898, 2003–04) paras 326–70; Iraq Inquiry, *The Report of the Iraq Inquiry: Executive Summary* (HC 264, 2016–17) paras 630–807; PHIA Common Analytical Standards (n 3).

[18] NIST SP 800-92 (n 8) 2-1–2-6; NIST SP 800-201 (n 1) 15–26; Council of Europe Convention on Cybercrime (opened for signature 23 November 2001, entered into force 1 July 2004) ETS No 185, art 16; Crime (Overseas Production Orders) Act 2019; National Cyber Security Centre, *Cyber Incident Response* https://www.ncsc.gov.uk/section/about-ncsc/incident-management accessed 6 August 2026.

[19] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) arts 6 and 8; *Big Brother Watch and Others v United Kingdom* (2022) 74 EHRR 17, [332]–[347]; *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Kärntner Landesregierung* (Joined Cases C-293/12 and C-594/12) EU:C:2014:238, [51]–[69]; *Tele2 Sverige AB v Post- och telestyrelsen and Secretary of State for the Home Department v Watson* (Joined Cases C-203/15 and C-698/15) EU:C:2016:970, [99]–[125]; *La Quadrature du Net and Others v Premier ministre and Others* (Joined Cases C-511/18, C-512/18 and C-520/18) EU:C:2020:791, [118]–[139]; Council of Europe Convention on Cybercrime (n 19) arts 16–18 and 23–35; Second Additional Protocol to the Convention on Cybercrime on Enhanced Co-operation and Disclosure of Electronic Evidence (opened for signature 12 May 2022) CETS No 224; Agreement between the Government of the United Kingdom of Great Britain and Northern Ireland and the Government of the United States of America on Access to Electronic Data for the Purpose of Countering Serious Crime (signed 3 October 2019, entered into force 3 October 2022) TS No 33/2024; *Microsoft Corporation v United States* 829 F 3d 197 (2d Cir 2016), vacated as moot 584 US 236 (2018); *R (KBR Inc) v Director of the Serious Fraud Office* [2021] UKSC 2, [2022] AC 519.

Preservation nevertheless engages privacy, confidentiality, privilege and foreign law. The European human-rights and data-retention cases reject indiscriminate or inadequately supervised acquisition merely because national security is invoked. The Budapest Convention, its Second Additional Protocol, overseas production legislation and the United Kingdom–United States Data Access Agreement provide routes for specified data, but none abolishes necessity, jurisdiction, provider rights or the protection of unrelated persons.[20]

The answer is granular preservation. The order identifies data class, account, system, time range, reason, owner, expiry and route to challenge. It preserves the smallest technically sufficient dependency and segregates uninvolved data. Where a provider cannot isolate the relevant material, the state should record why a broader image is necessary, apply access controls and revisit scope. A broad secret reservoir assembled ‘in case it becomes useful’ is not an attribution system; it is a different power requiring its own legal basis.

Correction must travel with the conclusion. Every recipient of an attribution assessment—minister, ally, regulator, provider, sanctions authority or prosecutor—should receive a material correction with the same priority as the original. A dashboard which updates silently is insufficient; the decision made upon the earlier version may already have produced consequence. MOSAIC’s correction propagation and append-only history therefore perform a constitutional function: they enable the State to act provisionally without pretending that provisional fact became immutable when first signed.

3.5 Numerical Confidence and Calibration

Numbers can discipline judgement where they describe a defined event and are tested against outcomes. They can also conceal it. ‘Eighty per cent confidence that State X is responsible’ may combine uncertainty concerning common operation, operator identity, contractual relationship, State status, legal control and the possibility that the visible infrastructure was compromised. The multiplication is rarely shown; the base rates are usually unknown; the audience receives a number whose precision cannot be reproduced.

Likelihood and confidence should therefore remain separate. Likelihood describes the assessor’s judgement that a proposition is true. Confidence describes the strength of the information and analytical process supporting that judgement. A proposition may be assessed as likely upon low-confidence reporting because little evidence exists; another as only a realistic possibility upon high-confidence evidence because two explanations remain genuinely balanced. PHIA’s standards expressly distinguish uncertainty, assumptions, information gaps and analytical confidence; ICD 203 requires the basis of uncertainty and source quality to be explained.[21]

Calibration is nevertheless valuable. An organisation which repeatedly uses ‘highly likely’ should examine whether comparable propositions were later borne out at approximately the intended rate, while recognising selection effects and the difficulty of obtaining ground truth. It should audit which indicator classes produced false positives, how often allied corroboration was independent, and whether particular teams systematically became more confident when the proposed response was politically preferred. The purpose is institutional learning, not conversion of State responsibility into a bookmaker’s market.

The legal threshold must not be encoded as the same probability across powers. Reasonable suspicion cannot sensibly be expressed as forty per cent, nor proof beyond reasonable doubt as ninety-five. Legal

[20] PHIA Common Analytical Standards (n 3); Intelligence Analysis, ‘Explaining Uncertainty’ (n 2); Office of the Director of National Intelligence (n 2) 3–6; Kent (n 2) 49–65.

[21] *Re B* (n 16) [2], [13] and [70]–[72]; *Woolmington* (n 5) 481–82; Crown Prosecution Service (n 5) paras 4.6–4.8; Judicial College, *Crown Court Compendium Part I: Jury and Trial Management and Summing Up* (October 2025) pt 5-1 https://www.judiciary.uk/wp-content/uploads/2022/06/Crown-Court-Compendium-Part-I.pdf accessed 6 August 2026.

standards allocate risk, prescribe institutional roles and determine what kinds of evidence may count; probability is only part of that work. *Re B* rejected a sliding civil standard tied to seriousness, whilst recognising that serious and improbable allegations may require stronger evidence before the ordinary balance is met. The criminal standard similarly concerns the tribunal's state of satisfaction upon admissible proof, not a legislated number.[22]

Decision-support software should consequently display proposition, standard, source quality and consequence separately. A red indicator beside 'high confidence' must not authorise a power automatically. The official remains required to choose the legal gateway, test the facts which it requires and record why the consequence of error is acceptable. Automation can ensure that a field is completed; it cannot decide that a secret inference is sufficiently compelling to freeze property or use force.

A numerical model is most useful below the legal conclusion: estimating time to data loss, propagation probability, likely service interruption, civilian dependency, cost of alternative containment and time required for a warrant or host response. Those estimates allow necessity and urgency to be tested. Their uncertainty should be shown by ranges and assumptions. A model which predicts catastrophic effect if access is used does not identify the actor who holds the access.

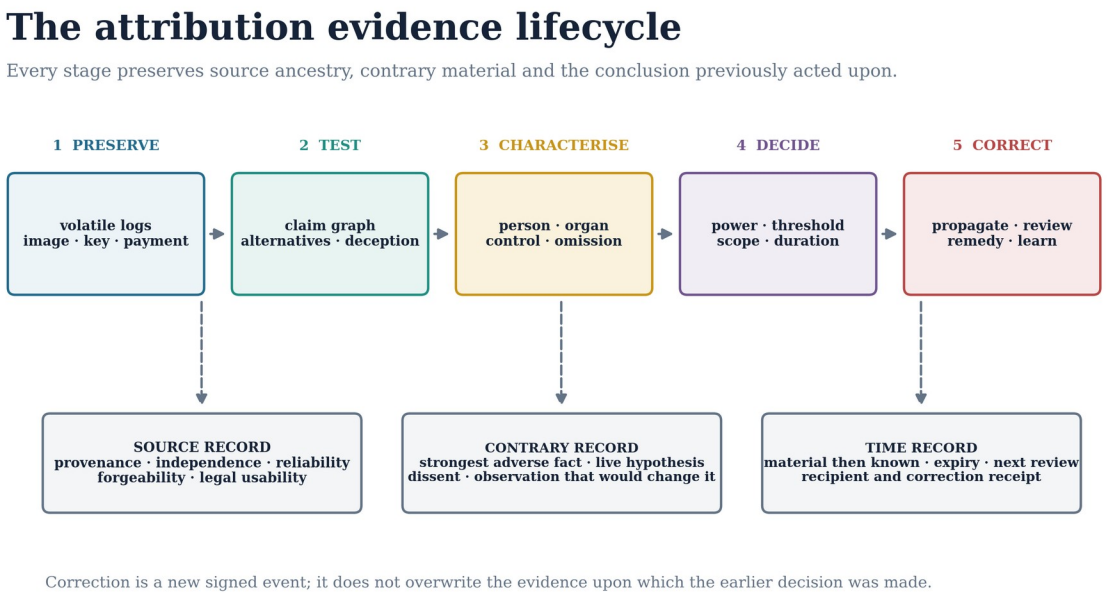

Figure 2. Evidence moves through preservation, hypothesis testing, legal characterisation, decision and correction; every stage retains the contrary material and prior state.

4. The Response-Specific Thresholds

4.1 Defensive Preservation and Containment

The lowest rung concerns acts whose immediate purpose is to preserve fact or prevent an identified system from completing harm, not to punish an author. It includes retaining volatile records, warning a victim, invalidating a compromised credential, isolating a malicious workload, blocking a command domain, rate-limiting traffic and suspending one transaction. Existing criminal, investigatory, data-

[22] Computer Misuse Act 1990 (n 10); Investigatory Powers Act 2016; Data Protection Act 2018; Network and Information Systems Regulations 2018, SI 2018/506; Cyber Security and Resilience (Network and Information Systems) Bill, HL Bill 32 of 2026–27 (as brought from the Commons, 17 June 2026); Department for Science, Innovation and Technology, 'Summary of the Bill' (30 June 2026) https://www.gov.uk/government/publications/cyber-security-and-resilience-network-and-information-systems-bill-factsheets/summary-of-the-bill accessed 6 August 2026.

protection and sectoral regimes supply parts of that authority; the Cyber Security and Resilience Bill's reporting and regulatory machinery strengthens the flow of incident information but does not create one general attribution power.[23]

The appropriate threshold is recorded, particularised reasonable grounds to believe that the object contains material evidence or is presently enabling the harmful operation, coupled with a reasonable belief that delay would materially prejudice preservation or prevention. Four safeguards do the constitutional work: the measure is confined to the smallest effective object; its duration is short; collateral dependencies are identified; and continuation requires independent confirmation. The threshold is lower because the act is provisional and reversible, not because cyber danger suspends law.

Containment should generally turn upon object attribution, not ultimate authorship. If authenticated telemetry shows that workload W is issuing destructive commands to the victim, government may not need to know whether its controller is a criminal group, PMSC or State organ before isolating W under a lawful power. The relevant proposition is that W is the operative dependency and isolation will probably reduce harm without disproportionate collateral effect. Demanding proof of State sponsorship at this rung would confuse the identity necessary for later coercion with the object necessary for defence.

Maximum credible harm properly has its greatest threshold effect here. If a credential opens an NHS data repository or safety controller, a reasonable possibility of destructive use may justify temporary suspension sooner than it would for an ordinary commercial account. The consequence affects urgency and acceptable temporary inconvenience. It does not authorise destruction of the account holder's property, public accusation or action against unrelated systems.

4.2 Confidential Diplomatic and Provider Engagement

The second rung is confidential engagement with a State, provider, ally or company capable of preserving, explaining or stopping the operation. Diplomacy frequently proceeds upon intelligence which could not lawfully sustain punishment. A State may ask another to investigate infrastructure, warn that continued activity risks escalation, or supply indicators through a protected channel without publishing its full assessment. OSCE confidence-building measures and the international framework of cyber points of contact exist in part because clarification and co-operation may be valuable before public legal positions harden.[24]

The threshold is an honest all-source assessment that the recipient probably possesses relevant control, information or capacity, even if ultimate responsibility remains unresolved. The communication must state what is being asked and avoid presenting an investigative request as a concluded accusation. Where urgency permits, it should distinguish: 'infrastructure in your jurisdiction is involved'; 'an entity subject to your authority appears to control it'; and 'we assess that your State directed the operation'. Those are different messages.

Confidentiality is not a licence to threaten upon invention. The record should identify the technical proposition, the confidence and the leading alternative. A deliberately false diplomatic attribution may itself violate international obligations or destroy the trust upon which later requests depend. But the law

[23] Organization for Security and Co-operation in Europe, Decision No 1202, 'OSCE Confidence-Building Measures to Reduce the Risks of Conflict Stemming from the Use of Information and Communication Technologies' (10 March 2016) PC.DEC/1202; UNGA, 'Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security: Final Substantive Report' (10 March 2021) UN Doc A/AC.290/2021/CRP.2, paras 28–31; Cybercrime Convention Committee, *24/7 Points of Contact Network of the Budapest Convention on Cybercrime: Assessment Report* (T-CY(2020)2, 2020); Council of Europe Convention on Cybercrime (n 19) art 35.

[24] Eichensehr (n 4) 570–98; Egloff, 'Public Attribution of Cyber Intrusions' (n 4); Chimène I Keitner, 'Attribution by Indictment' (2019) 113 *AJIL Unbound* 207, 207–12 https://doi.org/10.1017/aju.2019.34; Kristen E Eichensehr, 'Decentralized Cyberattack Attribution' (2019) 113 *AJIL Unbound* 213, 213–17 https://doi.org/10.1017/aju.2019.35.

should not require a government to publish intelligence before asking a host State to preserve a server. Such a rule would warn the controller, endanger sources and ensure that co-operation arrived after the evidence had gone.

4.3 Public Attribution

Public attribution is neither a judgment nor a legally neutral press release. It can impose reputational and economic cost, mobilise allies, expose intelligence capability, shape insurance and contractual decisions, identify private persons and prepare later coercive measures. It also performs an expressive function: attribution by indictment and coordinated government statements seek to declare which behaviour is unacceptable and to deny the attacker the benefit of anonymity.[25]

No general rule of international law requires a State publicly to disclose all evidence upon which its attribution rests. Several published State positions expressly preserve national discretion concerning the form and extent of evidence released. That absence of a disclosure duty does not mean that a public allegation requires no evidential discipline. The more specific and damaging the allegation, the stronger the corroboration and the more exact the language should be.[26]

The proposed threshold is corroborated all-source confidence sufficient for a reasonable decision maker to state the specified proposition publicly after considering source quality, independence, alternative explanations and foreseeable consequence. At least one evidential root should ordinarily be capable of public description, even if sensitive detail is withheld. Where nothing material can be shown, the statement should explain the class of evidence, the role of allies and the degree of confidence rather than feign proof through adjective.

Public attribution should name its level. ‘We assess that cluster C conducted the operation’ is technical. ‘We assess that company P supplied infrastructure and operators’ is organisational. ‘The United Kingdom holds State X responsible’ may be a policy conclusion; if it asserts legal attribution, the statement should identify, at least in general terms, the route—organ, empowered entity, instruction, control, adoption or assistance. A State may rationally decline that precision for intelligence reasons, but the internal certificate must contain it.

The March 2024 United Kingdom attribution concerning Chinese State-affiliated actors illustrates the value of date, scope and verb. The government distinguished the Electoral Commission compromise from reconnaissance against parliamentarians, identified an entity and two individuals for sanctions and coordinated the announcement with partners. The accompanying United States indictment carried far

[25] Foreign, Commonwealth & Development Office, ‘Application of International Law to States’ Conduct in Cyberspace: UK Statement’ (3 June 2021) https://www.gov.uk/government/publications/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement accessed 6 August 2026; Australian Government, *Australia’s Position on How International Law Applies to State Conduct in Cyberspace* (2020); France, Ministry of the Armies, *International Law Applied to Operations in Cyberspace* (2019); Government of the Netherlands, ‘Letter to the Parliament on the International Legal Order in Cyberspace’ (5 July 2019); Federal Government of Germany, ‘On the Application of International Law in Cyberspace’ (March 2021); Government of Canada, *International Law Applicable in Cyberspace* (22 April 2022); New Zealand Ministry of Foreign Affairs and Trade, *The Application of International Law to State Activity in Cyberspace* (1 December 2020); UNGA, ‘Official Compendium of Voluntary National Contributions on the Subject of How International Law Applies to the Use of Information and Communications Technologies by States’ (13 July 2021) UN Doc A/76/136.

[26] Foreign, Commonwealth & Development Office, ‘UK Holds China State-affiliated Organisations and Individuals Responsible for Malicious Cyber Activity’ (25 March 2024) https://www.gov.uk/government/news/uk-holds-china-state-affiliated-organisations-and-individuals-responsible-for-malicious-cyber-activity accessed 6 August 2026; National Cyber Security Centre, ‘UK Calls Out China State-affiliated Actors for Malicious Cyber Targeting of UK Democratic Institutions and Parliamentarians’ (25 March 2024) https://www.ncsc.gov.uk/news/china-state-affiliated-actors-target-uk-democratic-institutions-parliamentarians accessed 6 August 2026; US Department of Justice, ‘Seven Hackers Associated with Chinese Government Charged with Computer Intrusions Targeting Perceived Critics of China and US Businesses and Political Officials’ (25 March 2024) https://www.justice.gov/opa/pr/seven-hackers-associated-chinese-government-charged-computer-intrusions-targeting accessed 6 August 2026; Council of the European Union, ‘Cyber-attacks: Six Individuals Added to EU Sanctions List for Malicious Cyber Activities against EU Member States and Ukraine’ (24 June 2024).

more granular allegations because it served a criminal as well as diplomatic function. Neither document was a conviction. Each should be cited for what it was.[27]

Correction at this rung must be public when the error was public and material. A quiet amendment cannot repair a named person's reputation or an ally's policy formed upon the original. The correction should identify the proposition withdrawn, the reason so far as safely possible, the measures affected and whether the underlying threat assessment remains. Institutional embarrassment is not a lawful reason to leave a known falsehood in circulation.

4.4 Sanctions and Designation

Sanctions impose coercive legal consequence upon a named person and may affect property, travel, company control and ordinary life. Under the United Kingdom's post-2018 framework the applicable regulations ordinarily require reasonable grounds to suspect that the person is involved in or connected with specified activity, while the Act imposes procedural and review requirements. The Russia Regulations exemplify that structure. The relevant attribution is personal and statutory: it is not enough that an intrusion is broadly associated with a country or that the proposed designation may signal resolve.[28]

'Reasonable grounds' is not a minister's unexaminable assertion. There must be objective material capable of supporting the specified suspicion, and the statutory description must fit. The decision must pursue the sanctions purpose and withstand the proportionality and review required by the governing scheme. *Ahmed* demonstrates the constitutional danger of drastic asset-freezing power created without sufficient parliamentary authority; *Bank Mellat* insists upon rational connection and proportionality; *Youssef* addresses review in the international-listing context; and *Shvidler* confirms both the breadth of contemporary sanctions policy and the continuing judicial task.[29]

The threshold is therefore lower than proof on the balance of probabilities but higher than association by narrative. Payment, contract, beneficial ownership, tasking, operational reporting, repeated infrastructure support or knowing facilitation may establish reasonable grounds. Nationality, friendship, presence in the same industry or one shared service ordinarily cannot do so without the statutory connection. Where a designation rests materially upon classified evidence, the procedure must protect the person's effective ability to challenge the essence of the case so far as the governing law requires.

Urgent designation may be justified where funds or access will disappear, but urgency changes procedure and duration, not the facts which the statute requires. The initial decision record should identify the property or capability at risk, the evidential basis, the link to the sanctioned activity and the date for full review. An emergency label cannot repair a designation aimed at a person whom government believes dangerous but cannot connect to the criterion Parliament enacted.

Sanctions should likewise distinguish operator from payer. A technical report may establish that operator O ran the infrastructure. A bank record and contract may establish that company P paid O. Intelligence may indicate that a State service instructed P. A designation of P cannot borrow confidence

---

[27] Sanctions and Anti-Money Laundering Act 2018 (n 5) ss 11–12 and 23; Russia (Sanctions) (EU Exit) Regulations 2019 (n 5) reg 6; HM Treasury, *OFSI Guidance for the UK Financial Sanctions Regime* (December 2025) 20–35; Foreign, Commonwealth & Development Office, *Post-Legislative Scrutiny Memorandum: Sanctions and Anti-Money Laundering Act 2018* (CP 1020, March 2024) paras 102–19.

[28] *Ahmed v HM Treasury* [2010] UKSC 2, [2010] 2 AC 534, [34]–[61]; *Bank Mellat* (n 5) [68]–[76]; *Youssef v Secretary of State for Foreign and Commonwealth Affairs* [2016] UKSC 3, [2016] AC 1457, [30]–[38]; *Shvidler v Secretary of State for Foreign, Commonwealth and Development Affairs* [2025] UKSC 30, [57]–[93]; Sanctions and Anti-Money Laundering Act 2018 (n 5) ss 23–28.

[29] *American Cyanamid* (n 5) 406–09; *Ninemia Maritime Corp v Trave Schiffahrtsgesellschaft mbH (The Niedersachsen)* [1983] 1 WLR 1412 (CA) 1422; *Fourie v Le Roux* [2007] UKHL 1, [2007] 1 WLR 320, [25]–[30]; *Norwich Pharmacal Co v Customs and Excise Commissioners* [1974] AC 133 (HL) 174–75; *Bankers Trust Co v Shapira* [1980] 1 WLR 1274 (CA) 1282; Senior Courts Act 1981, s 37; Civil Procedure Rules 1998, pt 25.

from the attribution of O without proving the connecting edge. The claim graph prevents the political desire to ‘sanction the network’ from turning every node into the author of every act.

4.5 Civil Restraint, Disclosure and Asset Measures

Civil courts have long acted upon provisional evidence where waiting for trial would defeat justice. Interim injunctions, freezing orders and disclosure orders are conditioned by their own tests, undertakings, notice rules and equitable discipline. *American Cyanamid* directs attention to serious issue and balance; freezing jurisdiction requires a good arguable case and a real risk that judgment will go unsatisfied; *Norwich Pharmacal* and *Bankers Trust* supply distinct routes by which an innocent intermediary may be required to identify a wrongdoer or trace property.[30]

The cyber-economic application is immediate. A provider may hold the only subscriber record; an exchange the only route from ransom to wallet; a cloud company the only snapshot of an attacking instance. The court need not decide final authorship before preserving or disclosing that material. It must decide whether the legal test for the particular order is met, whether the respondent is the proper object, how innocent data will be protected and what undertaking or compensation should follow if the applicant was wrong.

Overseas data complicates but does not abolish the question. The Crime (Overseas Production Orders) Act, the Budapest system and the United Kingdom–United States Agreement provide defined routes. *Microsoft* exposed the territorial limits of the earlier United States statute; *KBR* rejected an unbounded reading of a domestic compulsory power against a foreign company. Legal process must follow the location, custody, treaty and statutory route actually available, not the fiction that data are nowhere because users experience the cloud as placeless.[31]

Civil restraint may properly rest upon a stronger provisional case than preservation but less than final liability. The decision maker should ask whether the respondent is sufficiently connected to the material or asset, whether the order is necessary to prevent frustration, whether a narrower object exists, and who bears loss if attribution changes. A secret executive database should not be permitted to perform the work of a court order merely because the subject is a foreign hacker.

4.6 Criminal Investigation, Charge and Conviction

Criminal law exposes most clearly why one attribution cannot serve every purpose. Arrest requires reasonable suspicion grounded in objective fact; it does not require evidence sufficient to convict. *O’Hara* and the Strasbourg authorities distinguish the information which can justify suspicion from proof at trial. Search, seizure, equipment interference and production powers each carry statutory conditions. The investigation may begin while identity remains one hypothesis, provided that coercion is tied to the lawful threshold and reasonable lines of inquiry include material pointing away from guilt.[32]

---

[30] Crime (Overseas Production Orders) Act 2019 (n 19); Council of Europe Convention on Cybercrime (n 19) arts 16–18 and 23–35; Second Additional Protocol (n 20); UK–US Data Access Agreement (n 20); *Microsoft* (n 20); *KBR* (n 20); Clarifying Lawful Overseas Use of Data Act 2018, Pub L No 115-141, div V, 132 Stat 1213.

[31] *O’Hara* (n 5) 298–99; *Fox, Campbell and Hartley v United Kingdom* (1990) 13 EHRR 157, [32]–[35]; Police and Criminal Evidence Act 1984 (n 5) ss 8, 19 and 24; Investigatory Powers Act 2016 (n 23) pts 5–6; Home Office, *Equipment Interference Code of Practice* (December 2022) paras 3.10–3.21.

[32] Crown Prosecution Service (n 5) paras 4.6–4.14 and 5.1–5.11; Crown Prosecution Service, *Director’s Guidance on Charging* (6th edn, December 2020, revised March 2025) paras 6.1–6.18 https://www.cps.gov.uk/prosecution-guidance/directors-guidance-charging-sixth-edition-december-2020-incorporating-national accessed 6 August 2026; Crown Prosecution Service, ‘Cybercrime—Prosecution Guidance’ https://www.cps.gov.uk/prosecution-guidance/cybercrime-prosecution-guidance accessed 6 August 2026.

Charge is different. The Code for Crown Prosecutors requires sufficient admissible, reliable and credible evidence to provide a realistic prospect of conviction against each suspect on each charge, followed by the public-interest stage. The Threshold Test permits an urgent charge only within defined conditions and where identifiable further evidence is expected; speculation that the cyber evidence may improve is not enough. The prosecutor must consider the likely defence, including compromise of the accused's system, shared credentials, false flag, lack of knowledge and an autonomous agent's departure from instructions.[33]

Conviction requires the tribunal to be sure. *Woolmington*'s allocation of the criminal burden remains the starting point, subject to lawful exceptions. Attribution by cluster, intelligence assessment or allied confidence cannot displace proof of the defendant's acts and mental state. A person may be a member of a contractor group without writing the destructive module; may write the module without knowing the target; may administer infrastructure without sharing the objective. Collective danger is not personal guilt.

Disclosure is part of the evidential threshold, not an inconvenience following it. The Criminal Procedure and Investigations Act and *R v H* require unused material capable of undermining the prosecution or assisting the defence to be handled through a process compatible with a fair trial. Intelligence cannot be converted into admissible proof by withholding the material which reveals its weakness. If necessary protection makes a fair prosecution impossible, government may use another lawful response; it may not secure conviction by trying a different case in secret.[34]

United States practice of attribution by indictment illustrates both strength and limitation. Detailed indictments can expose actors, accounts, infrastructure, victims and State employment with a granularity absent from diplomatic statements; they may support arrest if a defendant travels and can coordinate allies. They remain allegations until plea or judgment. Keitner's analysis is valuable because it shows that indictment is simultaneously coercive, deterrent and expressive whilst still governed by criminal evidential expectations.[35]

4.7 Countermeasures

A countermeasure is an act which would otherwise breach an international obligation of the injured State, taken against the responsible State to induce compliance with its obligations of cessation and reparation. It is not punishment, does not include force, must be directed against the responsible State, remains subject to proportionality and ordinarily requires a prior call for compliance and notice. The Articles on State Responsibility express those limits; *Gabčíkovo-Nagymaros* and the *Air Services Agreement* supply important foundations in practice.[36]

The necessary attribution is therefore exacting in a way that public condemnation need not be. The decision maker must identify the prior internationally wrongful act, the obligation breached, the injured

[33] Criminal Procedure and Investigations Act 1996 (n 13) ss 3 and 7A; *R v H* (n 13) [13]–[36]; ECHR (n 20) art 6; *Edwards and Lewis v United Kingdom* (2005) 40 EHRR 24, [46]–[55].

[34] Keitner (n 25) 207–12; US Department of Justice, 'Six Russian GRU Officers Charged in Connection with Worldwide Deployment of Destructive Malware and Other Disruptive Actions in Cyberspace' (19 October 2020) https://www.justice.gov/archives/opa/pr/six-russian-gru-officers-charged-connection-worldwide-deployment-destructive-malware-and accessed 6 August 2026; US Department of Justice, 'Seven Hackers Associated with Chinese Government' (n 27); US Department of Justice, 'Justice Department Charges 12 Chinese Contract Hackers and Law Enforcement Officers in Global Computer Intrusion Campaigns' (5 March 2025) https://www.justice.gov/opa/pr/justice-department-charges-12-chinese-contract-hackers-and-law-enforcement-officers-global accessed 6 August 2026.

[35] International Law Commission (n 5) arts 49–54; *Gabčíkovo-Nagymaros Project (Hungary/Slovakia)* [1997] ICJ Rep 7, [82]–[87]; *Air Services Agreement (United States of America v France)* (Award) (1978) 18 RIAA 417, [81]–[98]; Crawford (n 10) 685–99.

[36] UK cyber statement (n 26); Australian Government (n 26); Government of the Netherlands (n 26); France, Ministry of the Armies (n 26); Michael N Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd edn, CUP 2017) rr 20–25 and commentary.

State, the conduct attributed to the proposed target State and the legal route of attribution. It is not sufficient that infrastructure lay in that State, that its citizens probably operated the campaign or that it failed to assist as quickly as desired. A countermeasure responding to an omission must identify the primary duty, knowledge or constructive knowledge, capacity and reasonable measures which the territorial State failed to take. An act of the private controller does not become the host State's act merely because an omission is proved.

International law has not adopted one numerical standard for a State's decision to take countermeasures. Published cyber positions generally insist upon a case-by-case all-source assessment, and several reject any legal obligation to disclose the underlying evidence publicly. That latitude does not remove the risk of wrongfulness: if the predicate breach or attribution is absent, the act remains a breach by the responding State. The legal consequence of error supplies the practical reason for a high and corroborated degree of confidence proportionate to the countermeasure's gravity, intrusiveness and reversibility.[37]

The standard should have four limbs. First, the facts necessary to establish breach and attribution must be supported by independent or demonstrably reliable material after deception alternatives are considered. Secondly, the legal characterisation must identify the rule and attribution mode rather than rely upon 'State-sponsored'. Thirdly, the proposed non-performance must be capable of inducing compliance and must not affect prohibited obligations or third-party rights. Fourthly, the certificate must state the cessation condition and the evidence upon which the measure will be suspended or terminated.

Collective political action should not be mislabelled. Allies may coordinate retorsion—unfriendly but lawful acts—upon a lower evidential risk because no international obligation is suspended. Whether States other than the injured State may take collective countermeasures in the general interest remains contested. The United Kingdom may coordinate sanctions, procurement, diplomatic and defensive acts which are independently lawful without resolving that controversy. The internal record should identify which measure is retorsion and which depends upon the law of countermeasures.[38]

An autonomous or sharded operation does not relax the predicate. If the evidence proves that a State paid a PMSC for the result, supplied target classes and received operational reporting, those facts may support instructions, direction or control depending upon their relationship to the conduct. If the evidence proves only a general contract for 'strategic pressure', with the agent selecting every unlawful means, the case may establish assistance, procurement offences or a failure of due diligence without necessarily satisfying article 8. The desire to impose cost cannot choose the attribution rule after the evidence is known.

4.8 Force in Self-Defence

Force is the terminal rung because error is least reversible and the legal conditions are most demanding. Article 51 preserves self-defence if an armed attack occurs; State practice and scholarship address imminence where an attack is actual or genuinely imminent. Cyber means may constitute an armed attack where their scale and effects are equivalent, including reasonably anticipated death, injury or

[37] International Law Commission (n 5) art 54 and commentary; Martin Dawidowicz, *Third-Party Countermeasures in International Law* (CUP 2017) 111–73; Crawford (n 10) 703–11; UK cyber statement (n 26).

[38] UN Charter (n 5) arts 2(4) and 51; Schmitt (n 37) rr 69–71 and commentary; Oona A Hathaway and others, 'The Law of Cyber-Attack' (2012) 100 *California Law Review* 817, 836–58; Heather Harrison Dinniss, *Cyber Warfare and the Laws of War* (CUP 2012) 73–113; Marco Roscini, *Cyber Operations and the Use of Force in International Law* (OUP 2014) 69–118; Nicholas Tsagourias, 'Cyber Attacks, Self-Defence and the Problem of Attribution' (2012) 17 *Journal of Conflict and Security Law* 229, 233–57.

physical destruction. The classification of the operation, the identity of the attacker, necessity, target and proportionality are separate propositions.[39]

The burden cannot be discharged by saying that intelligence is secret or that the victim must act. In *Oil Platforms* the Court placed upon the State invoking self-defence the burden of showing an armed attack attributable to the State against which force was used, and examined necessity and proportionality. *Nicaragua* and *Armed Activities* likewise demonstrate that assistance, frontier activity and support do not automatically establish an armed attack by the State which force is proposed against. Whatever debates attend attacks by non-State actors, the response must be directed at a lawful target and must satisfy the territorial and necessity questions applicable to the facts.[40]

The proposed domestic decision standard is a high degree of confidence in every proposition whose error would make the use of force unlawful or direct it at the wrong target. That normally requires corroboration across independent evidential roots or an exceptionally reliable direct source; express consideration of deception and innocent control; legal certification; target-specific current intelligence; and a record of the feasible non-forcible alternatives. ‘High degree’ is not a claim that international law has enacted a percentage. It is the minimum constitutional discipline appropriate to an act capable of killing upon a secret executive record. The formulation is a proposal about domestic decision discipline rather than a claim that international law has enacted a verbal standard, and the reason it has to be a proposal is itself established in the literature. Green has shown that the Court has applied noticeably different evidentiary formulations across Nicaragua, Oil Platforms and Armed Activities without reconciling them, and Ruys reaches a comparable conclusion from the practice on armed attack. A State acting in self-defence therefore cannot derive its standard from the jurisprudence, and the choice to require a high degree of confidence for propositions whose error would make the force unlawful is defended here on the ground of irreversibility rather than presented as a rule already in force.[41]

Imminence does not permit the evidence to be vague. It changes time. The Attorney General’s published principles and Bethlehem’s formulation examine whether the threat is actual or imminent, whether further evidence can be obtained without losing the opportunity to act and whether the expected attack is part of a continuing pattern. A destructive payload already staged in a winter grid may demand a decision before execution; that temporal fact can make waiting unreasonable. It cannot prove who staged it.[42]

Maximum credible harm is relevant to the armed-attack assessment where the operation’s reasonably anticipated consequences form part of scale and effects. Erasure of clinical data, disabling of safety systems or prolonged winter power loss may cause death without explosive force. The assessment must be grounded in access, demonstrated capability, target dependence, persistence, timing and feasible

[39] *Oil Platforms* (n 15) [51]–[78]; *Nicaragua* (n 11) [191], [195] and [199]; *Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v Uganda)* (Judgment) [2005] ICJ Rep 168, [143]–[147]; Olivier Corten, *The Law Against War* (2nd edn, Hart 2021) 487–532.

[40] Attorney General’s Office, ‘Attorney General’s Speech at the International Institute for Strategic Studies’ (11 January 2017) https://www.gov.uk/government/speeches/attorney-generals-speech-at-the-international-institute-for-strategic-studies accessed 6 August 2026; Daniel Bethlehem, ‘Self-Defense against an Imminent or Actual Armed Attack by Nonstate Actors’ (2012) 106 *American Journal of International Law* 770, 775–77; Christopher Greenwood, ‘International Law and the Pre-emptive Use of Force: Afghanistan, Al-Qaida, and Iraq’ (2003) 4 *San Diego International Law Journal* 7, 13–23.

[41] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3, arts 48, 51(3), 51(5)(b), 52(2) and 57; ICRC, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (Nils Melzer ed, ICRC 2009) 65–73; Ministry of Defence, *The Manual of the Law of Armed Conflict* (OUP 2004) paras 5.4–5.33; ICRC, *International Humanitarian Law and Cyber Operations during Armed Conflicts* (Position Paper, November 2019) 4–8; Schmitt (n 37) rr 92–120 and commentary.

[42] UK cyber statement (n 26); Attorney General’s Office, ‘Cyber and International Law in the 21st Century’ (23 May 2018) https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century accessed 6 August 2026; UNGA Official Compendium (n 26); Bethlehem (n 42) 775–77; Corten (n 40) 553–91.

propagation. A speculative worst case cannot turn a completed theft into an armed attack, and a severe threat in general cannot identify one person as the live controller.

Necessity requires more than proof of authorship. If a provider can disable the credential, a host State can arrest the operator in time, or a reversible cyber act can stop the payload with materially lower collateral risk, force may not be necessary. Conversely, the mere theoretical existence of a slower alternative does not defeat necessity where the evidential clock shows that it cannot act before the threatened effect. The certificate should state the estimated time to harm, the time and probability associated with each alternative, and the evidence for both.

Targeting law then performs work which attribution alone cannot. Civilians and civilian objects remain protected unless and for such time as the applicable rules remove that protection; military objectives are defined by nature, location, purpose or use and the definite military advantage offered by destruction, capture or neutralisation. Precautions and proportionality apply. A person who wrote code months earlier is not necessarily the person now operating the attack; a server used for command may share civilian functions; an autonomous system may have migrated. Live-target verification is therefore a continuing duty until effect.[43]

Where doubt concerns only whether the attack was commissioned by a State, but reliable evidence establishes an actual or imminent armed attack by a non-State actor and identifies the target capable of stopping it, the law concerning self-defence against non-State actors and territorial sovereignty must be addressed honestly. The United Kingdom's published position accepts that self-defence may be available where the host State is unwilling or unable to prevent the attack, a proposition not universally accepted. The disagreement must appear in the certificate. It cannot be concealed by attributing the non-State attack to the host upon lesser evidence.[44]

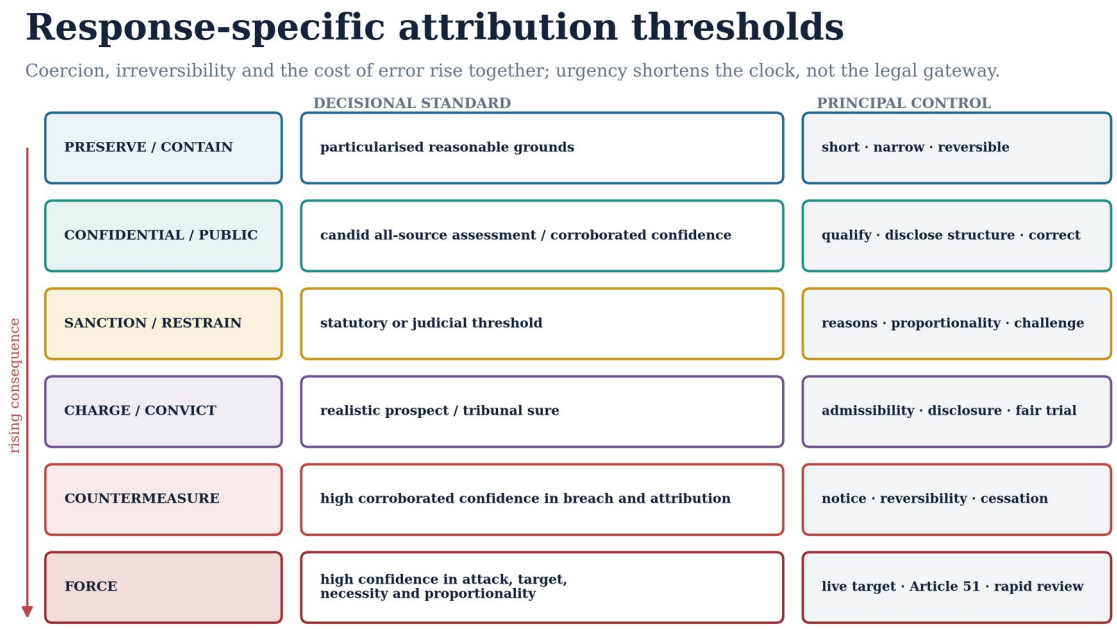

Figure 3. The attribution threshold rises with coercion, irreversibility and the consequence of error; urgency accelerates review but does not erase the gateway.

[43] International Law Commission (n 5) arts 4 and 7 and commentaries; Crawford (n 10) 113–26; *Difference Relating to Immunity from Legal Process of a Special Rapporteur of the Commission on Human Rights* (Advisory Opinion) [1999] ICJ Rep 62, [62].

[44] International Law Commission (n 5) art 8 and commentary; *Nicaragua* (n 11) [109]–[115]; *Bosnian Genocide* (n 11) [396]–[407]; *Prosecutor v Tadić* (Appeal Judgment) IT-94-1-A (15 July 1999), [116]–[145]; Crawford (n 10) 141–56; Aust (n 10) 191–249.

5. Attribution to States, Auxiliaries and Territorial Hosts

5.1 Organs and Empowered Entities

The easiest legal route is sometimes the hardest to observe. Conduct of any State organ is attributable under article 4 whatever the organ's function and position, and ultra vires conduct remains attributable where the organ acts in that capacity. A military cyber unit which uses commercial hosting, personal accounts and foreign infrastructure does not become private. Proof of organ status and operational conduct is sufficient; proof that the cabinet ordered the particular intrusion is not additionally required.[45]

Article 5 reaches an entity which is not formally an organ but is empowered by domestic law to exercise elements of governmental authority and acts in that capacity. The category is narrower than 'important contractor'. A company selling intrusion capability, satellite imagery or cloud services does not exercise governmental authority merely because government is its principal customer. The assessment must identify the legal empowerment, governmental function and capacity in which the impugned act occurred.

Those rules reveal why technical and legal attribution can move in opposite directions. Investigators may be uncertain which individual typed the command but know, from authenticated internal infrastructure and status evidence, that the command came from an organ. Conversely, they may identify the natural person beyond serious dispute whilst lacking evidence that his private side work was performed in official capacity. The attribution certificate records both.

5.2 Instructions, Direction and Control

Article 8 concerns private persons or groups acting upon a State's instructions or under its direction or control. *Nicaragua* and the *Bosnian Genocide* judgment apply an effective-control approach to attribution of the particular conduct; the *Tadić* overall-control test was developed for a different classification question and should not be imported without explanation. Contract, payment, training, shared objectives and equipment may be powerful facts, but they do not each answer whether the State instructed, directed or controlled the operation complained of.[46]

The paper's difficult case is deliberate functional delegation. A State service contracts with a PMSC to disable a rival's insurance and municipal systems. The PMSC integrates an agent and subcontractors. The State specifies protected exclusions, target classes, budget, duration and the strategic outcome, receives daily campaign metrics and can terminate payment; it does not approve each exploit. The evidence may support instruction at the level of a campaign even where the machine selects the individual technique and victim. Control should not be defined so microscopically that delegation to autonomy becomes a legal solvent.

Yet the opposite error is equally serious. A State which pays for intelligence, buys a vulnerability or tolerates a contractor's unrelated crime has not necessarily instructed the resulting attack. The certificate should identify: the operative instruction; its level of specificity; the controller's retained discretion; the State's stop or modification power; reporting and benefit; and the connection between those facts and the act alleged to be wrongful. Where the evidence supports procurement or assistance but not article 8, the law should say so.

[45] International Law Commission (n 5) art 11 and commentary; *Tehran Hostages* (n 12) [58]–[74]; Crawford (n 10) 166–69.

[46] International Law Commission (n 5) art 16 and commentary; Aust (n 10) 191–249; *Bosnian Genocide* (n 11) [420]–[422].

5.3 Acknowledgment, Adoption and Assistance

A State may acknowledge and adopt private conduct as its own under article 11. *Tehran Hostages* illustrates that approval and maintenance after the initial private seizure could convert the legal position. Mere political praise, refusal to condemn or exploitation of an outcome is not necessarily adoption. In cyber affairs, a State's later publication of stolen material, operational tasking of the same access or express acceptance of the campaign may be relevant; silence ordinarily is not.[47]

Article 16 separately concerns a State which knowingly aids or assists another State in committing an internationally wrongful act, where the conditions of the article are met. It should not be used to describe every company or individual accomplice, for the Articles govern responsibility between States. The facts—financing, infrastructure, intelligence, credentials, sanctuary—may also engage domestic offences, sanctions or due diligence. The legal route must remain attached to the actor whose responsibility is asserted.[48]

The importance is practical. A countermeasure against State A as principal requires attribution of the prior act to A. A sanction against company P may require only statutory involvement. Diplomatic action against State B for assistance may depend upon article 16. A defensive order against server S may depend only upon its present function. One operation can support all four responses upon different propositions and evidence.

5.4 Territorial Knowledge and the Attribution of Omissions

Evelyne Schmid and Ayşe Özge Erceiş demonstrate with particular clarity that attribution of a State's omission is conceptually distinct from attributing the attacker's conduct. Due diligence asks whether the territorial or otherwise relevant State possessed the requisite knowledge and capacity and failed to take reasonable measures required by the applicable primary rule. The omission is the State's own. It does not turn a private attacker into an organ.[49]

That distinction becomes indispensable in a sharded campaign. State A hosts one encrypted inference; State B a short-lived relay; State C a payment intermediary; State D the controller. No State except D may see target, objective or campaign. It would be perverse to assemble the global picture after the event and impute it retrospectively to every territorial State. The victim must prove the knowledge and feasible protective action relevant to each omission, not the maximum knowledge of the network as a whole.

Corfu Channel permits inference where facts lie within a State's exclusive control, but it does not create strict liability for everything crossing its territory. Pulp Mills and modern due-diligence authorities relate obligation to risk, knowledge, capacity and reasonable conduct. Published cyber positions differ concerning whether customary international law contains a specific cyber due-diligence obligation.

---

[47] Schmid and Erceiş (n 1) 581–90; Antonio Coco and Talita de Souza Dias, 'Cyber Due Diligence: A Patchwork of Protective Obligations in International Law' (2021) 32 *European Journal of International Law* 771, 777–805; Russell Buchan, 'Cyberspace, Non-State Actors and the Obligation to Prevent Transboundary Harm' (2016) 21 *Journal of Conflict and Security Law* 429, 433–55; Joanna Kulesza, *Due Diligence in International Law* (Brill Nijhoff 2016) 247–66.

[48] *Corfu Channel* (n 15) 18 and 22; *Pulp Mills* (n 16) [197]; UK cyber statement (n 26); France, Ministry of the Armies (n 26); Australian Government (n 26); Government of the Netherlands (n 26); Michael N Schmitt, 'In Defense of Due Diligence in Cyberspace' (2015) 125 *Yale Law Journal Forum* 68, 68–81; Eric Talbot Jensen and Sean Watts, 'A Cyber Duty of Due Diligence: Gentle Civilizer or Crude Destabilizer?' (2017) 95 *Texas Law Review* 1555, 1559–93.

[49] Swiss Federal Department of Foreign Affairs and ICRC, *Montreux Document on Pertinent International Legal Obligations and Good Practices for States related to Operations of Private Military and Security Companies during Armed Conflict* (17 September 2008); Foreign, Commonwealth & Development Office, *The Pall Mall Process Code of Practice for States* (4 April 2025) https://www.gov.uk/government/publications/the-pall-mall-process-code-of-practice-for-states accessed 6 August 2026; Maurer (n 1) 27–50; Deborah D Avant, *The Market for Force: The Consequences of Privatizing Security* (CUP 2005) 1–18.

That disagreement itself forbids the easy assertion that failure to stop one foreign workload is automatically an internationally wrongful act.[50]

Notice changes the future position. An authenticated and particular request which identifies infrastructure, victim, time and the reasonable step sought may give a State knowledge it previously lacked. Refusal, delay or assistance to migration may then support a separate inference concerning omission, and in an extreme case complicity. The notice must be precise enough to act upon; a mass list of foreign addresses accompanied by a threat is not proof that each host can lawfully or technically disable them.

The same principle protects weak States and improves co-operation. If providing logs or isolating an account risks later being treated as confession of responsibility, States will assist less. A doctrine which distinguishes innocent fragment, negligent omission and controlled campaign permits the victim to demand action without telling every host that co-operation proves guilt.

5.5 PMSCs, Contractors and Autonomous Agents

PMSCs and commercial cyber operators occupy the gap between private crime and organ. The Montreux Document restates obligations and good practice relating to PMSCs without making every contractor act attributable to the client State. The Pall Mall Process addresses commercial intrusion capability and State practice but is not an attribution code. Maurer's account of cyber proxies shows why States value the distance which such actors provide.[51]

The evidential model should identify payer, prime, subcontractor, integrator, deployer, operator and infrastructure controller. It should also identify which actor supplied objective, target constraints, tools, access, budget, stop authority and acceptance of result. A legal system which asks only who pressed the key will fail where the machine selected the command; one which asks only who paid will punish innocent finance. Functional allocation supplies the facts from which the applicable responsibility rule can operate.

Open-source or locally hosted AI sharpens this need. Once weights, tools and credentials are placed under the contractor's control, the model provider may neither observe nor revoke use. The agent can choose targets and routes within an objective. Responsibility does not migrate into the machine. It remains with the persons and organisations whose conduct satisfies the relevant rule: the State which instructed, the integrator which intentionally deployed, the operator which continued after warning, or the host which knowingly omitted a required and feasible measure.[52]

[50] National Cyber Security Centre, *The Near-term Impact of AI on the Cyber Threat* (n 9) 4–9; Google Threat Intelligence Group, *Adversarial Misuse of Generative AI* (January 2025) 4–28 https://cloud.google.com/resources/content/adversarial-misuse-of-generative-ai accessed 6 August 2026; NIST AI RMF 1.0 (n 14) 17–31; Guidelines for Secure AI System Development (n 14); NIST SP 800-218 (n 14).

[51] Rid and Buchanan (n 2) 14–23; Buchanan (n 17) 31–64; Egloff, 'Public Attribution of Cyber Intrusions' (n 4); US Department of Justice, 'US Government Disrupts Botnet People's Republic of China Used to Conceal Hacking of Critical Infrastructure' (31 January 2024) https://www.justice.gov/opa/pr/us-government-disrupts-botnet-peoples-republic-china-used-conceal-hacking-critical accessed 6 August 2026; Cybersecurity and Infrastructure Security Agency and others, 'Enhanced Visibility and Hardening Guidance for Communications Infrastructure' (3 December 2024) https://www.cisa.gov/resources-tools/resources/enhanced-visibility-and-hardening-guidance-communications-infrastructure accessed 6 August 2026.

[52] UN Charter (n 5) art 51; Schmitt (n 37) rr 69–71 and commentary; Roscini (n 39) 69–118; Dinniss (n 39) 73–113; *Oil Platforms* (n 15) [51]–[78]; *Nicaragua* (n 11) [191]–[195].

Autonomy may, however, affect proof of a particular mental element. An engineer who wrote a general scheduler need not know the attack. A prime which deliberately prevents subcontractors from seeing the whole may retain the only culpable integration. A controller cannot rely upon ignorance which it designed for others as proof that it lacked the full objective. The manifest proposed in Papers 2–5 is therefore an evidential allocation device: it records the human and organisational decisions which would otherwise disappear behind machine selection.

6. False Flags, Campaigns and the Designed Appearance of Another

6.1 False Flags Are Operations, Not Magic

A false flag may plant language artefacts, reuse public malware, imitate working hours, route through a rival's infrastructure, steal code-signing material, seed a political message or manufacture a payment trail. It may also exploit the defender's expectations: investigators who believe a particular State uses a tool may stop testing alternatives when the tool appears. Imperfect attribution creates a strategic opportunity because an attacker can seek both impunity and escalation between others.[53]

The correct response is not permanent agnosticism. False flags have costs and leave evidence. The attacker requires access to the imitated tool or artefact, knowledge of the rival's practice, infrastructure and timing consistent with the deception, and operational discipline sufficient to preserve it. Some indicators are cheap to copy; others—long-lived access, non-public victim knowledge, internal tasking, authenticated payment, human sources and infrastructure observed across years—are materially harder. The assessment should rank indicators by forgeability and control rather than count them.

Technical deception and legal deception also differ. An operator may convincingly imitate State X's malware but cannot as easily fabricate a bank's authenticated record showing that State Y paid the prime contractor, unless it has compromised that record too. Conversely, payment by Y may be genuine whilst the contractor exceeds its instructions. The claim graph permits each edge to be attacked separately, preventing one discovered false flag from erasing independent evidence and preventing one genuine link from validating planted technical detail.

6.2 Campaign Aggregation

Cyber operations are often classified too narrowly when each intrusion is assessed alone and too broadly when every event sharing a label is treated as one campaign. Aggregation matters to intent, scale, effects, imminence, sanctions involvement and whether an isolated mistake is in fact repeated policy. It also risks circular proof: events are attributed to a campaign because they resemble it, and the campaign is attributed to a State because it contains those events.

The certificate should state the aggregation rule. Common control may be inferred from authenticated infrastructure management, shared non-public tooling, coordinated target selection, temporal hand-off, common payment or operational reporting. Mere similarity of victim sector, publicly available exploit

[53] Foreign & Commonwealth Office and National Cyber Security Centre, 'Foreign Office Minister Condemns Russia for NotPetya Attacks' (15 February 2018) https://www.gov.uk/government/news/foreign-office-minister-condemns-russia-for-notpetya-attacks accessed 6 August 2026; US Department of Justice, 'Six Russian GRU Officers Charged' (n 35); Cybersecurity and Infrastructure Security Agency, 'Advanced Persistent Threat Compromise of Government Agencies, Critical Infrastructure, and Private Sector Organizations' (AA20-352A, updated 15 April 2021) https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-352a accessed 6 August 2026; Council of the European Union, 'Russian Cyber Operations against Ukraine: Declaration by the High Representative on behalf of the European Union' (10 May 2022) https://www.consilium.europa.eu/en/press/press-releases/2022/05/10/russian-cyber-operations-against-ukraine-declaration-by-the-high-representative-on-behalf-of-the-european-union/ accessed 6 August 2026; FCDO APT31 statement (n 27).

or political rhetoric is weaker. Where events are aggregated for the harm assessment but not for actor attribution, that difference must be explicit.

International-law classification likewise requires care. A series of operations may be relevant to scale and effects or to an imminent continuing attack, but the law concerning accumulation of events is contested and context-specific. It cannot be used to add unrelated low-confidence incidents until the armed-attack threshold appears numerically satisfied. Each event's connection to the campaign and actor must be evidenced before its effect is added.[54]

6.3 Four Public Attribution Records

Public records demonstrate differing evidential forms. The United Kingdom and partners attributed NotPetya to the Russian Government in 2018; later United States charges supplied individual and organisational allegations concerning GRU officers. The 2021 SolarWinds statements combined government attribution, sanctions and detailed technical advisories. In 2022 the European Union and partners attributed the Viasat operation to Russia in the context of the invasion of Ukraine. In 2024 the United Kingdom's APT31 announcement combined a domestic democratic-system incident, reconnaissance against parliamentarians, sanctions and coordinated allied action.[55]

These records should not be flattened. A technical advisory may contain indicators and mitigations without stating the whole intelligence case. A sanctions announcement applies a domestic criterion. An indictment pleads facts which remain to be proved. A joint diplomatic statement expresses political confidence and collective consequence. Their agreement can be powerful where based upon independent national collection; it is not automatically independent merely because several States publish on the same day.

Volt Typhoon shows another form. The February 2024 joint advisory described persistent access to United States critical infrastructure, pre-positioning and living-off-the-land techniques, and attributed the activity to PRC State-sponsored actors. Its value for defensive action lay not only in the State label but in the detailed behaviour, affected sectors and mitigations which operators could test. A victim can contain upon the technical proposition even if it cannot see the intelligence supporting the governmental relationship.[56]

The lesson is not that governments must publish sources which would be destroyed by publication. It is that every public product should disclose the proposition it is equipped to support. Technical detail supports detection and common-operation inference; legal instruments support the selected coercive consequence; intelligence supports propositions which may have to remain described rather than exposed. A glossy joint statement cannot substitute for the internal map connecting them.

[54] NCSC and others, 'PRC State-Sponsored Actors' (n 7); US Department of Justice, 'US Government Disrupts Botnet' (n 53); Cybersecurity and Infrastructure Security Agency and others, *Primary Mitigations to Reduce Cyber Threats to Operational Technology* (6 May 2025) https://www.cisa.gov/resources-tools/resources/primary-mitigations-reduce-cyber-threats-operational-technology accessed 6 August 2026.

[55] PHIA Common Analytical Standards (n 3); Office of the Director of National Intelligence (n 2) 3–6; Review of Intelligence on Weapons of Mass Destruction (n 18) paras 326–70; Iraq Inquiry (n 18) paras 630–807.

[56] *Al Rawi v Security Service* [2011] UKSC 34, [2012] 1 AC 531, [22]–[48]; Justice and Security Act 2013, pt 2; Special Immigration Appeals Commission Act 1997; Special Immigration Appeals Commission (Procedure) Rules 2003, SI 2003/1034; Regulation of Investigatory Powers Act 2000, s 65.

Four public attribution products

Each product supports a different proposition. Agreement does not convert one form into another.

PRODUCT	PRIMARY PROPOSITION	EVIDENTIAL FORM	WHAT IT DOES NOT PROVE
technical advisory	behaviour · infrastructure mitigation	telemetry · indicators reproducible method	State control or personal guilt
diplomatic statement	government assessment and intended consequence	all-source intelligence qualified public reasons	criminal conviction or every source
designation	named person meets statutory criterion	ministerial record reviewable reasons	international attribution of the whole campaign
indictment	probable criminal case against named defendants	pleaded facts · counts and admissible-evidence plan	guilt before plea or judgment

Publication rule: state the actor level, confidence, classes of evidence, legal consequence and whether criminal allegations remain unproved.

Figure 4. Public attribution products carry different propositions and burdens: advisory, diplomatic statement, designation and indictment must not be treated as interchangeable proof.

7. Intelligence, Closed Material and Allied Confidence

7.1 Intelligence May Decide Without Becoming Trial Evidence

Government may lawfully make foreign-policy, defensive and national-security decisions upon intelligence which would not be admissible at trial. Intelligence can include intercept, agent reporting, allied liaison, covert access and analytic inference whose collection or disclosure engages separate law. Requiring every ministerial decision to rest upon open evidence would expose sources and make timely defence impossible. Treating the same material as sufficient for every legal consequence would be equally destructive.

The first rule is purpose control. The certificate states whether material is used to generate a hypothesis, establish reasonable grounds, support an executive assessment or prove a fact in proceedings. It records restrictions upon disclosure and use. An ally’s caveated report may legitimately increase confidence for a confidential warning whilst being unavailable for public attribution; it should not be cited internally as ‘independent corroboration’ if the recipient cannot test its source ancestry.

The second rule is source-conscious confidence. PHIA and ICD 203 standards require quality, access, credibility, age, denial and deception, alternatives and confidence to be addressed. A minister should receive not merely the conclusion but a source summary identifying which propositions depend upon single-source reporting, which are independently corroborated and which rest upon assumption. Policy urgency is not a reason to remove the caveat most likely to change the decision.[57]

The third rule is separation between secret reason and public account. Government may decline to publish a source. It should not publish a different factual basis known to be insufficient because the real basis is secret. A bounded statement—‘our assessment draws upon technical, intelligence and allied

[57] *A and Others v United Kingdom* (2009) 49 EHRR 29, [205]–[220]; *Tariq v Home Office* [2011] UKSC 35, [2012] 1 AC 452, [27]–[45]; *Secretary of State for the Home Department v AF (No 3)* [2009] UKHL 28, [2010] 2 AC 269, [57]–[65]; *Bank Mellat v HM Treasury (No 1)* [2013] UKSC 38, [2014] AC 700, [28]–[35].

reporting which cannot be disclosed in full'—is more honest than presenting a forgeable indicator as though it alone proved State direction.

7.2 Closed Material Requires Law

English law does not recognise a general executive entitlement to prove coercive action through material which the affected person and ordinary lawyer cannot see. *Al Rawi* held that closed material procedure in ordinary civil litigation required statutory authority. Parliament supplied a defined regime in Part 2 of the Justice and Security Act 2013. Specialised contexts, including SIAC and the Investigatory Powers Tribunal, have their own bases.[58]

Even where closed procedure is lawful, fairness does not disappear. *A and Others*, *Tariq* and the control-order cases examine whether the affected person receives sufficient information to instruct the special advocate and answer the essence of the case, with the requirement depending upon the legal context and consequence. The special advocate mitigates secrecy; he does not make undisclosed evidence equivalent to evidence fully tested by the person who knows the facts.[59]

Haralambous demonstrates that a closed procedure may be implied where necessary to make a statutory search-warrant and judicial-review scheme effective, but it does not create a general power. Public-interest immunity remains different: material withheld on PII grounds is ordinarily withheld from court as well as the other party, whereas closed material is considered by court and special advocate. *Conway v Rimmer* and *Wiley* require the public interest in secrecy to be balanced within that doctrine.[60]

Paper 5's designated court must therefore receive express authority for defined protective, designation and review proceedings if closed material is to be used. The statute should require open and closed reasons, the minimum secrecy compatible with the interest protected, special advocates where appropriate, regular review of classification and an obligation to provide the gist necessary for effective challenge to the greatest extent lawful. It should not authorise a criminal conviction upon a closed case.

7.3 Allied Intelligence and the Risk of Circular Confidence

Collective attribution can strengthen evidence, legitimacy and deterrence. Allies possess different collection access, victim telemetry, languages and jurisdictional powers. Joint assessment can expose error which one service would miss. Yet liaison also creates circular reporting. Agency A's judgement may pass to B, return through C and appear to A as independent confirmation. A common technical vendor may underlie every national report.

The claim graph should record evidential roots across allies without requiring broad disclosure of source identity. Each partner can attest whether its conclusion depends materially upon information received from another participant, whether it possesses an independent collection root and what proposition that root supports. A 'five allies agree' statement then has meaning beyond arithmetic.

Caveats must survive political co-ordination. An ally may permit use for defence but not sanctions, or confidential decision but not publication. It may assess the operator with confidence but not the State

[58] *R (Haralambous) v Crown Court at St Albans* [2018] UKSC 1, [2018] AC 236, [42]–[60]; *Conway v Rimmer* [1968] AC 910 (HL) 952–53; *R v Chief Constable of the West Midlands Police, ex p Wiley* [1995] 1 AC 274 (HL) 306–07; *Al Rawi* (n 58) [22]–[48].

[59] Egloff and Smeets (n 4) 502–33; Eichensehr (n 4) 570–98; Office of the Director of National Intelligence, *Intelligence Community Directive 208: Maximizing the Utility of Analytic Products* (9 January 2017) 1–3 https://www.dni.gov/files/documents/ICD/ICD-208-Maximizing-the-Utility-of-Analytic-Products-2017-01-09.pdf accessed 6 August 2026; UK cyber statement (n 26).

[60] FCDO APT31 statement (n 27); NCSC APT31 statement (n 27); US Department of Justice, 'Seven Hackers Associated with Chinese Government' (n 27); Council of the European Union, 'Cyber-attacks: Six Individuals Added' (n 27).

relationship. The United Kingdom cannot cure that limit by paraphrasing the report into its own assessment without acknowledging the dependence. Nor should one ally possess an unexplained veto over correction once contrary evidence emerges.

Alliance does not transfer legal responsibility for decision. The minister who designates, the prosecutor who charges and the State which takes a countermeasure must be satisfied under the law governing that act. Reliance upon allied material may be reasonable; ‘our partners were confident’ is not a statutory criterion.

7.4 The Public Case Without Source Sacrifice

A democratic State cannot always publish the material which convinced it. It can usually publish more about the structure of its conclusion than the bare formula ‘we assess’. The public case should state the incident and time period, the conduct attributed, the level of actor identified, the degree and basis of confidence, the classes of evidence used, material technical indicators safe to disclose, allied participation, the legal or policy consequence selected and whether proceedings remain allegations.[61]

The APT31 announcement and allied materials show several available layers: an NCSC account directed to democratic institutions and parliamentarians; a governmental responsibility and sanctions statement; and a United States indictment naming persons, companies, accounts and alleged acts. The products protected undisclosed intelligence whilst still allowing the public to distinguish the Electoral Commission compromise, reconnaissance and criminal allegations. Precision of scope did not require publication of every source.[62]

Public reasons are particularly important where private companies contribute attribution. A vendor may have excellent visibility and a commercial interest in publishing a threat report; it may also use proprietary methods, customer telemetry and cluster labels which cannot be independently reconstructed. Government should identify whether it verified the underlying proposition, relied upon a vendor report as one source or merely adopted its nomenclature. Outsourcing collection must not become outsourcing public responsibility.

A minimum public case also improves correction. If government states only a conclusion, a later correction appears to reverse an indivisible belief. If it states propositions, it can explain that infrastructure and operator findings remain while the State-control inference has changed. The public can then assess whether continuing defensive measures and terminated coercive measures fit the corrected evidence.

There are limits. Revealing a non-public vulnerability, covert access or human source may extinguish future collection or endanger life. A court may accept closed material within an authorised procedure. Parliament may receive a classified supplement. The obligation proposed here is not maximal disclosure; it is maximal specificity compatible with the protected interest, paired with an internal certificate containing what public reasons omit. *Al Rawi*’s insistence that fundamental procedural change requires law and *A and Others*’ concern with the essence of the case warn against allowing administrative convenience to become a complete answer.[63]

---

[61] *Al Rawi* (n 58) [22]–[48]; *A and Others* (n 59) [205]–[220]; Justice and Security Act 2013 (n 58) pt 2; Freedom of Information Act 2000, ss 23–24.

[62] *Corfu Channel* (n 15) 18 and 22; *Bosnian Genocide* (n 11) [209] and [401]–[407]; *Wisniewski v Central Manchester Health Authority* [1998] PIQR P324 (CA) P340; *Keefe v Isle of Man Steam Packet Co Ltd* [2010] EWCA Civ 683, [16]–[21]; *Armory v Delamirie* (1722) 1 Strange 505, 93 ER 664.

[63] ISO/IEC 27037:2012 (n 8); NIST SP 800-92 (n 8); SLSA, ‘Supply-chain Levels for Software Artifacts Specification v1.0’ (2023) https://slsa.dev/spec/v1.0/ accessed 6 August 2026; NIST SP 800-218 (n 14); National Telecommunications and Information Administration, *The Minimum Elements for a Software Bill of Materials* (12 July 2021) 5–15.

Public confidence is not strengthened by pretending certainty. It is strengthened when government states what it attributes, distinguishes assessment from charge and judgment, acknowledges material limits and corrects error. An adversary will exploit caveat rhetorically; that cost is smaller than teaching allies, courts and citizens that official attribution is language which cannot be examined and need never be repaired.

8. Fragmentation, Adverse Inference and the Controller's Evidential Burden

8.1 Designed Opacity

An attacker may distribute code, data, payment and execution so that each provider and territorial State sees an innocent fragment. The controller retains the map, yet structures contracts and permissions to ensure that no subcontractor does. If ordinary evidential rules then treat every fragment in isolation, design has converted opacity into exculpation. The law need not accept that result.

Adverse inference is a rule of evaluation, not a substitute offence. Domestic courts may draw appropriate inferences from a party's failure to call evidence or preserve material where the circumstances justify it; international courts have likewise considered exclusive control and non-production when evaluating proof. *Corfu Channel* permits more liberal inference where direct evidence lies within the territorial State's control, while *Bosnian Genocide* insists that exceptional gravity remains matched by fully conclusive proof. The two propositions coexist: obstruction affects what may reasonably be inferred, but does not abolish the burden or the substantive rule.[64]

The proposed controller inference requires five conditions. First, the person had possession or control of specified material which would probably resolve a material attribution proposition. Secondly, a lawful and particular preservation or production duty applied. Thirdly, the person knew of the duty and had a reasonable opportunity to comply. Fourthly, the material was withheld, destroyed or fragmented without adequate explanation. Fifthly, the inference drawn is no wider than the proposition which the missing material could reasonably prove.

Thus a PMSC which deletes the tasking record after an authenticated order may face an inference that the record was adverse upon the scope of tasking. The inference does not prove that the client State exercised effective control over every operation. A cloud provider which never possessed plaintext target data cannot be treated as though it concealed them. A weak territorial State which retained one router log cannot bear the controller's missing campaign map.

8.2 No Transfer of Ignorance to the Territorial State

The central discipline of Paper 3 remains. A victim State's later global picture must not be projected backwards into each host's knowledge. Notice, capacity, access and reasonable action are assessed locally and temporally. The fact that MOSAIC correlates ten jurisdictions proves why federation is needed; it does not prove that each State was culpable before correlation.

Where a host receives a particular notice and refuses even to preserve the identified record, the future assessment changes. Knowledge may be proved by the notice; capacity by the provider and law; omission by the refusal. If the host warns the controller or assists migration, stronger conclusions may follow. The duty remains a duty concerning the host's own conduct.

[64] *Council of Civil Service Unions v Minister for the Civil Service* [1985] AC 374 (HL) 410–11; *Bank Mellat* (n 5) [68]–[76]; *Shvidler* (n 29) [57]–[93]; *Kennedy v Charity Commission* [2014] UKSC 20, [2015] AC 455, [51]–[56].

This limitation is not softness. It concentrates consequence upon the actor who designed the ignorance. A payer, prime or integrator which possesses objective, target policy, manifests and reports can be subjected to preservation, licensing, sanctions, criminal and civil duties within the applicable jurisdiction. Evidence should follow control. Treating an innocent host as the attacker because the controller escaped is both unjust and strategically foolish.

8.3 The Manifest as Evidence, Not Immunity

An attribution manifest records objective owner, model and tool versions, target constraints, credentials, infrastructure, subcontractors, changes, stop authority, material outputs and known incidents. It does not certify legality and should not create a safe harbour merely because boxes were completed. Its purpose is to ensure that a controller capable of deploying strategic private power cannot later claim that no record exists of the decisions by which it was constituted.

The manifest should be signed, append-only, access-controlled and preserved outside the runtime capable of altering it. Cryptographic integrity can show that a record has not changed since signature; it cannot show that the entry was true. Independent logs, payment records, provider receipts and human approvals remain necessary. The evidential system must not confuse tamper evidence with substantive accuracy.[65]

Good-faith completion and prompt self-reporting should matter. They reduce uncertainty, assist containment and may mitigate regulatory penalty. Deliberate falsification, suppression or deployment without the required manifest should aggravate the inference and offence. The aim is not paperwork for its own sake. It is to allocate the cost of designed opacity to the person who chose it.

9. Error, Correction and Compensation

9.1 Error Is a Foreseeable State, Not an Institutional Scandal

Any attribution system which promises no error is either dishonest or inactive. Adversaries conceal, providers lose records, allies disagree and urgent decisions precede complete evidence. The constitutional question is not whether government can guarantee omniscience, but whether it distinguishes reasonable provisional error from reckless or dishonest assertion, detects change and repairs consequence.

The first protection is contemporaneous record. Review should examine the evidence and alternatives available when the act was taken, not condemn every urgent decision because later forensics became better. Equally, a decision reasonable at issue may become unlawful in continuation when its basis erodes. Automatic expiry forces government to decide again; it prevents the administrative convenience by which a temporary suspicion acquires permanence through silence.

The second protection is proposition-specific correction. If infrastructure attribution remains sound but the payer link fails, containment may continue while designation ends. If the operator was correct but State instruction is disproved, prosecution may continue while a countermeasure ceases. The claim graph avoids the all-or-nothing instinct to defend every earlier conclusion lest one correction appear to destroy the whole case.

The third is independent review. Judicial review, statutory appeal, ministerial reconsideration, commissioner audit and parliamentary scrutiny perform different functions. *CCSU* confirms that national-security subject matter does not wholly exclude review, even where the court accords the

[65] *Factory at Chorzów (Germany v Poland)* (Merits) PCIJ Series A No 17, 47; International Law Commission (n 5) arts 30–31 and 34–39; Human Rights Act 1998, s 8; Crown Proceedings Act 1947.

executive appropriate latitude upon matters within institutional competence. *Bank Mellat* and *Shvidler* show that severe economic measures remain subject to legality and proportionality within the applicable statutory scheme.[66]

9.2 The Correction Duty

A material correction duty should arise where new information would probably have altered the attribution proposition, threshold assessment, measure, scope or duration. The duty bearer is the authority which issued or maintains the decision. It must notify operational recipients immediately; notify courts, regulators, prosecutors and allies through the appropriate protected route; and publish correction where the erroneous proposition was public, subject only to necessary protection of sources.

Correction must identify downstream consequence. A sanctions listing, watchlist, provider block, procurement exclusion and intelligence profile may each have copied the original. Removing the source record without tracing those copies preserves the harm while improving only the issuing authority's database. MOSAIC's recipient log should therefore be treated as part of remedy.

There is no legitimate 'confidence laundering' by citation. If Authority B adopted Authority A's assessment, A's correction must reach B even if B now describes the conclusion as its own. B may retain the conclusion upon independent evidence, but it should record that new basis and perform its own review. Allied political inconvenience cannot make a disproved source remain probative.

9.3 Remedy and Compensation

The form of remedy follows the act. A preservation order made reasonably but unnecessarily may require deletion and costs. An unlawful designation may require quashing, public correction and compensation where the governing law permits. Wrongful restraint may engage property, privacy, reputation and commercial loss. An unlawful international act carries the duties of cessation and full reparation expressed in *Factory at Chorzów* and the Articles on State Responsibility.[67]

Human-rights law adds procedural and substantive duties. Interference with private life, communications and possessions must possess a legal basis, pursue a legitimate aim and satisfy the applicable necessity and proportionality requirements. Effective challenge matters most where the State's case is secret. The Human Rights Act supplies domestic remedies within its field, but a bespoke statute should not force every correction into years of general litigation.[68]

Compensation should not be automatic for every reasonable emergency act, or officials will delay preservation until the loss is irretrievable. Nor should public necessity place all error upon an innocent person. Paper 5's model properly distinguishes good-faith temporary protective measures, for which defined compensation may follow exceptional loss, from reckless, dishonest or continued action after the basis failed. The decision certificate supplies the evidence necessary to administer that distinction.

---

[66] ECHR (n 20) arts 6, 8 and 10 and Protocol No 1, art 1; Human Rights Act 1998 (n 67) ss 3, 6 and 8; *Big Brother Watch* (n 20) [332]–[347]; *A and Others* (n 59) [205]–[220].

[67] Comptroller and Auditor General, *Investigation: WannaCry Cyber Attack and the NHS* (HC 2017–19, 414) 5–13; NHS England, 'Synnovis Cyber Incident' https://www.england.nhs.uk/synnovis-cyber-incident/ accessed 6 August 2026; Department of Health and Social Care, 'Synnovis Cyber Attack' (Written Statement HCWS1046, 30 June 2025).

[68] NCSC and others, 'PRC State-Sponsored Actors' (n 7); CISA and others, *Primary Mitigations* (n 56); Department for Energy Security and Net Zero, *National Emergency Plan 2023: Downstream Gas and Electricity* (July 2023) https://www.gov.uk/government/publications/national-emergency-plan-downstream-gas-and-electricity-2016 accessed 6 August 2026; Cybersecurity and Infrastructure Security Agency, *MAR-17-352-01 HatMan—Safety System Targeted Malware* (12 April 2018).

Personal consequence remains necessary for abuse. An official who honestly chooses the wrong hypothesis upon a complete record is not equivalent to one who suppresses dissent, fabricates confidence or continues a measure to avoid admitting error. Disciplinary, civil and, where the elements are met, criminal responsibility should turn upon the conduct, not merely the outcome. A system which punishes every error will conceal error; one which punishes none will repeat it.

10. Operational Tests

10.1 Destructive Access to NHS Data

An autonomous process obtains privileged access to a hospital group's records and begins testing bulk-deletion permissions. The command infrastructure rotates through two cloud providers; the payer is unknown; one model output contains a target list. The maximum reasonably credible harm is grave because loss of clinical data can delay diagnosis, treatment and transfusion, as the WannaCry and Synnovis records show in different ways.[69]

At 02.00 the lawful proposition is object-specific: the credential and workloads are enabling imminent destructive action. Particularised reasonable grounds justify suspension, preservation and victim warning. Ultimate authorship is unnecessary. The order records why waiting for payer identity would sacrifice the protective opportunity, isolates the smallest dependencies and preserves unrelated tenant data.

At 08.00 investigators correlate tool use with a known criminal cluster and a contractor account. That may justify confidential provider and host-State engagement and a public technical warning. It does not yet justify naming the contractor's home State as author. A later payment record and authenticated tasking channel may support designation of the contractor upon the statutory test. Criminal charge must still prove the named defendant's participation and knowledge through admissible evidence.

Force would remain unavailable merely because the harm could be catastrophic. The decision maker must prove an actual or imminent armed attack at the requisite scale and effects, identify the live target, establish necessity and satisfy the remaining law. If provider isolation stops the process, necessity ends. Maximum harm accelerated containment; it did not authorise escalation after containment succeeded.

10.2 Winter Energy Pre-positioning

Joint telemetry identifies living-off-the-land access maintained for months in energy-control environments. The activity resembles official descriptions of Volt Typhoon: persistent access and pre-positioning, using compromised infrastructure and ordinary tools to reduce obvious signatures. No destructive command has yet issued. Winter load means that coordinated interruption could endanger life.[70]

The immediate threshold supports evidence preservation, credential rotation, segmentation and, where authorised, isolation of confirmed hostile access. It also supports urgent engagement with providers and allies. Public attribution of the State relationship requires the all-source governmental assessment and

[69] US Department of Health and Human Services, Office for Civil Rights, 'Cyberattack on Change Healthcare' (Dear Colleague Letter, 13 March 2024) https://www.hhs.gov/sites/default/files/cyberattack-change-healthcare.pdf accessed 6 August 2026; UnitedHealth Group, *Change Healthcare Cyber Response Update* (7 May 2024) https://www.unitedhealthgroup.com/ns/changehealthcare.html accessed 6 August 2026; US Senate Committee on Finance, *Hacking America's Health Care: Assessing the Change Healthcare Cyber Attack and What's Next* (1 May 2024).

[70] US Department of Justice, 'Justice Department Charges 12 Chinese Contract Hackers and Law Enforcement Officers' (n 35); *United States v Yin Kecheng and Others*, Indictment, Case No 1:24-cr-00323 (DDC, filed 6 November 2024, unsealed 5 March 2025).

should state the proposition and confidence separately from the technical mitigations which operators can verify.

Imminence for force is not established merely by persistence plus winter. The certificate must assess target, capability, operational preparation, likely window, adversary decision and whether further evidence can safely be gathered. A dormant foothold may be espionage, contingency preparation or the beginning of attack. The inability to know with certainty does not forbid defence of the network; it does forbid treating every foothold as a licence to strike.

If intelligence later identifies an authorised campaign plan and live preparations for execution, the time-to-harm analysis changes. Non-forcible alternatives—revocation, isolation, host action and defensive cyber effect—must be tested in the time actually available. The more dangerous the payload, the stronger the duty to keep target verification current: destructive action against an abandoned controller neither prevents harm nor remains necessary.

10.3 Theft from an Insurer

A group steals policyholder data and demands payment. There is no evidence of danger to life or essential service. Failure to respond may encourage repetition and finance a wider campaign; the act is serious economic aggression and may be part of organised State procurement. Its gravity does not change the classification into an armed attack.

The response ladder nevertheless has force. Providers preserve records; exchanges restrain traced proceeds under lawful order; investigators pursue accounts and accomplices; public attribution warns victims; sanctions and procurement exclusions may reach a knowing payer; criminal process may follow. The Change Healthcare incident in the United States demonstrates that disruption in an insurance and payment intermediary can propagate across healthcare even where the immediate mechanism is not physical violence.[71]

If evidence identifies a foreign State as purchaser of repeated contractor operations, coordinated retorsion and sanctions may impose cost. A countermeasure requires the prior internationally wrongful act and attribution to that State. Force cannot be used to punish the completed theft or to create general deterrence. Firmness consists in selecting a measure which law permits and which reaches the dependency the actor values, not in choosing the most violent item upon the ladder.

10.4 The Foreign-Paid PMSC

A State ministry pays a PMSC to ‘reduce State V’s capacity to sustain sanctions’. The prime divides work among reconnaissance, access, model integration and infrastructure subcontractors. An agent selects local authorities and water utilities, reports aggregate disruption and migrates when blocked. The ministry receives daily progress and can modify target exclusions.

Technical attribution first connects workloads and campaign. Contract and payment identify the prime and payer. The reporting interface and modification power bear upon instruction or control. The legal assessment should not insist upon proof that the ministry approved each victim if the evidence establishes that the campaign complained of was the instructed object and the agent’s selection occurred within it. Nor should it attribute every subcontractor act to the State if the act fell outside that campaign.

[71] PHIA Common Analytical Standards (n 3); Review of Intelligence on Weapons of Mass Destruction (n 18) paras 326–70; Iraq Inquiry (n 18) paras 630–807; Intelligence and Security Committee of Parliament, *Russia* (HC 632, 2019–21) paras 28–50; Cabinet Office, *Professional Standards for Government Intelligence Assessment* (November 2018).

The operator and payer then face different thresholds. A short containment order follows the live workload. A designation follows reasonable grounds connecting the person to the specified activity. A criminal case against the integrator requires proof of intentional participation and relevant offence. Countermeasures against the payer State require the international breach and attribution route. Force requires the independent Article 51 and targeting conditions. The same evidence file feeds each; no decision borrows the gateway of another.

The March 2025 United States allegations concerning Chinese contract hackers and law-enforcement officers demonstrate why contractor structure is not hypothetical, while remaining charges rather than judgments. They support the need for an evidential model capable of connecting public officials, private companies, individual operators and campaigns without calling allegation conviction.[72]

10.5 The Fragmented AI System

A private billionaire pays a holding company which retains a PMSC. One coder in State A writes an optimiser; another in B a browser agent; a third in C a credential module. An integrator in D connects them to an open-weight model and stolen access. The objective is to force a foreign government to reverse policy by disrupting insurers, public services and communications. No coder except the integrator sees the whole.

The law should resist the theatrical question whether ‘the AI’ is responsible. The system is evidence of delegated selection. Attribution asks who defined objective and constraints, intentionally integrated capability, deployed it, monitored it, could stop it and financed it with knowledge. The innocent coder’s ignorance is relevant to his liability; it is not the integrator’s defence. The payer’s general hostility is not enough; contract, reporting, risk acceptance and control matter.

Territorial States A–C may have no knowledge that their fragments form an attack. Notice may create preservation and co-operation duties prospectively. State D, the controller’s State or the payer’s State may hold the decisive evidence. An adverse inference may follow if the integrator destroys the manifest after lawful notice. It cannot fill a missing article 8 relationship between payer and State.

The attribution certificate keeps the response honest. It may authorise containment of the deployed system while the payer remains unknown; civil disclosure to trace the holding company; sanction when the statutory connection is established; prosecution when admissible proof reaches the Full Code Test; and international response only upon the corresponding State-responsibility proposition. Autonomy makes the investigation messier. It does not make the gateways optional.

11. The Attribution Decision Certificate

11.1 Content

Every serious response should be accompanied by a certificate proportionate to urgency and consequence. The shortest emergency form may be completed in minutes; it must not be reconstructed after controversy begins. The certificate records:

1. the incident, protected interest and time;

[72] Nicholas Tsagourias and Michael Farrell, ‘Cyber Attribution: Technical and Legal Approaches and Challenges’ (2020) 31 European Journal of International Law 941, 941–67; Marco Roscini, ‘Evidentiary Issues in International Disputes Related to State Responsibility for Cyber Operations’ (2015) 50 Texas International Law Journal 233, 233–73; Martha Finnemore and Duncan B Hollis, ‘Beyond Naming and Shaming: Accusations and International Law in Cybersecurity’ (2020) 31 European Journal of International Law 969, 969–1003; Yuval Shany and Michael N Schmitt, ‘An International Attribution Mechanism for Hostile Cyber Operations’ (2020) 96 International Law Studies 196, 196–222; Kubo Mačák, ‘Decoding Article 8 of the International Law Commission’s Articles on State Responsibility: Attribution of Cyber Operations by Non-State Actors’ (2016) 21 Journal of Conflict and Security Law 405.

2. the precise technical, organisational and legal propositions asserted;
3. the measure proposed and the legal power relied upon;
4. the evidential threshold governing that measure;
5. supporting material, source ancestry and chain of custody;
6. material against the conclusion and live alternatives;
7. deception opportunity and the indicators least susceptible to imitation;
8. confidence in each material proposition and why;
9. actual and maximum reasonably credible foreseeable harm;
10. urgency, reversibility, collateral dependency and consequence of error;
11. narrower measures considered and the reason each is insufficient;
12. the decision owner, legal adviser, technical assessor and recorded dissent;
13. duration, review time, termination condition and live-target requirement;
14. recipients, disclosure classification and allied caveats; and
15. correction, notification, remedy and compensation routes.

The form is not a substitute for reasons. Its purpose is to force the reasons into the categories which error most often collapses. A minister can disagree with the technical assessor or accept risk which an analyst would not; the record shows that this was a policy decision rather than a scientific conclusion. The lawyer certifies the gateway, not the wisdom of the policy. The analyst assesses the proposition, not the lawfulness of force.

11.2 Threshold Matrix

The operative matrix is as follows:

Response	Proposition which must be established	Minimum decisional standard	Principal control
preservation	specified material is probably relevant and at risk	particularised reasonable grounds	narrow scope; short expiry; judicial or independent confirmation
defensive containment	identified object is enabling or imminently enabling harm	particularised reasonable grounds plus necessity	smallest effective dependency; collateral test; continuous review
confidential engagement	recipient probably has relevant knowledge, control or capacity	candid all-source assessment	qualify accusation; protect sources; record request and reply
public attribution	the stated actor relationship is sufficiently corroborated to assert publicly	corroborated, source-conscious confidence	proposition-specific language; public basis where feasible; correction duty
sanctions/designation	person satisfies the statutory connection to specified activity	governing statutory threshold, ordinarily reasonable grounds to suspect	reasons; proportionality; review; effective challenge
civil restraint/disclosure	claim and risk satisfy the order’s statutory or	governing judicial standard	court; undertaking; data minimisation;

Response	Proposition which must be established	Minimum decisional standard	Principal control
	equitable test		compensation
criminal charge	each offence has admissible, reliable and credible evidence making conviction more likely than not	Full Code Test, or lawful Threshold Test temporarily	independent prosecutor; disclosure; fair trial
conviction	defendant committed the offence and mental element	tribunal sure	open adversarial trial subject to lawful exceptions
countermeasure	prior breach is attributable to target State and measure satisfies international conditions	high, corroborated confidence proportionate to consequence of error	legal certificate; notice; proportionality; reversibility; cessation
force	actual or imminent armed attack, lawful target, necessity and proportionality	high degree of confidence in every dispositive factual proposition	law-officer and operational certification; live target; Article 51 report; rapid parliamentary review

The matrix deliberately refuses an arithmetical scale. A seventy per cent estimate can conceal a strong operator proposition and weak State-control proposition; a ninety per cent estimate says nothing about admissibility or necessity. The standard derives from the legal act, while confidence describes the assessment’s basis. Both must appear.

11.3 The Two Error Axes

Every certificate should compare false-positive and false-negative harm. A false positive may isolate an innocent tenant, freeze an innocent person, accuse an ally or kill the wrong target. A false negative may permit destruction of hospital data, grid failure or repeated attack. The balance changes with reversibility and time. That is why low-threshold preservation can be lawful where low-threshold force cannot.

The comparison must not be performed at the level of slogans. It should estimate who bears each error, the duration, the ability to compensate, the effect upon the threatened system and the likelihood that delay improves evidence. A one-hour credential suspension with compensation differs from indefinite designation; a reversible sinkhole differs from destruction of a shared server; a warning differs from indictment.

The decision owner also records the cost of waiting. Evidence may improve, remain static or disappear. Risk may grow, move or be contained by the victim. A higher threshold which can probably be reached within ten minutes is different from one dependent upon a foreign warrant expected in six weeks. Speed becomes a reasoned variable rather than an incantation.

11.4 Review Does Not Repeat the First Decision

The first reviewer asks whether the threshold was met when the measure issued and whether continuation is met now. The evidence may differ. He tests correction propagation, collateral effect,

compliance by the recipient, changed threat and whether a less intrusive route has become effective. Continuation is a new act, not an administrative echo.

For serious executive measures the reviewer should receive the rejected alternatives and dissent, not merely the final assessment. For closed material he should see source summaries and caveats. For force he must receive live target and necessity updates. For sanctions he must test current statutory connection and purpose. For prosecution the independent prosecutor applies the Code rather than inheriting the intelligence service's confidence label.

An audit trail which merely records that each official clicked 'approve' proves procedure, not judgment. The certificate must show the proposition each official was competent and responsible to decide. Distributed expertise then becomes a safeguard instead of a diffusion of responsibility.

11.5 Ownership and Institutional Dissent

No committee should own attribution in the abstract. The technical lead owns the integrity of the forensic proposition; the intelligence chair owns the all-source assessment and its caveats; the legal adviser owns the identification of the power and international gateway; the minister or other statutory officer owns the selected executive consequence. Each may depend upon the others, but none may sign for a judgement outside his competence and later say that 'government' decided it.

The institutional record should preserve dissent at the level at which it matters. A provider may accept common infrastructure but reject common control. An intelligence service may accept the contractor link but doubt State instruction. A lawyer may accept attribution but reject necessity for the proposed countermeasure. Compressing those disagreements into one confidence label ensures that the decision maker never sees the question which could alter the response.

Co-ordination should seek clarity rather than unanimity. PHIA standards require significant analytical differences to be communicated where concurrence cannot be achieved; the Iraq and Butler inquiries show why caveat must not disappear as an assessment approaches the policy customer.[73] The certificate should consequently state majority and dissenting positions, the evidence upon which they turn and whether the measure remains lawful upon the narrower view.

Responsibility continues after signature. The technical lead must report material telemetry change; the intelligence owner must propagate changed confidence; the lawyer must identify a gateway which has ceased to be available; the operational commander must terminate an effect which is no longer necessary; and the decision owner must ensure correction and remedy. A named owner is not an ornament placed upon the first page. It is the person who must act when the fact changes at 03.00 and the original committee has dispersed.

This allocation also makes speed possible. A pre-designated duty chain can preserve within minutes, contain within the authorised scope and convene the higher-level decision as the measure becomes more coercive. The alternative—discovering after the incident who may decide, who may see allied material and who can instruct a provider—does not protect liberty. It merely spends the period in which a narrow measure could have prevented the need for a severe one.

12. Conclusion

Attribution is often described as the problem which makes cyber response impossible. That description is convenient to both extremes: those who prefer inaction can demand certainty which no hostile

[73] James A Green, 'Fluctuating Evidentiary Standards for Self-Defence in the International Court of Justice' (2009) 58 International and Comparative Law Quarterly 163, 163–79; Tom Ruys, *'Armed Attack' and Article 51 of the UN Charter: Evolutions in Customary Law and Practice* (CUP 2010).

operation will supply, while those who prefer unconstrained power can say that uncertainty makes ordinary law obsolete. Neither proposition is sound.

The State already acts upon different evidential standards. It preserves upon reasonable grounds, investigates upon suspicion, prosecutes upon a realistic prospect, convicts only when the tribunal is sure, sanctions upon enacted criteria, restrains property through judicial tests, takes countermeasures upon a proved international predicate and may use force only within Article 51 and the remaining law. Cyber operations make the evidence distributed, volatile and deceptive; they do not merge those powers.

Technical attribution should state events, infrastructure, control and campaign. Legal attribution should state person, company, organ, empowerment, instruction, control, adoption, assistance or omission. Decisional attribution should state why the evidence satisfies the gateway for the act now proposed. The three may be completed at different times and by different institutions. Their separation is the foundation of speed because it permits government to preserve and contain without pretending prematurely to have proved payer, State and target.

The greatest danger does not lower the proof of identity. It changes urgency, preservation, precaution and the cost of temporary defensive error. Maximum reasonably credible foreseeable harm may justify isolating the smallest dependency before the destructive command is issued. It cannot convert suspicion of one State into a lawful countermeasure against it, still less turn a guessed location into a lawful target. Firmness consists in acting as soon as the applicable proposition is established and stopping as soon as it is not.

Classified and allied intelligence can support executive decisions, but secrecy is not probative weight. Closed process requires legal authority and an effective challenge appropriate to the consequence. Joint attribution requires source ancestry if agreement is to count as corroboration. Designed fragmentation can justify a bounded adverse inference against the controller which destroyed or withheld the map; it cannot project global knowledge into an innocent territorial State.

The Attribution Decision Certificate makes those conclusions operational. It records proposition, source, contrary material, rule, threshold, urgency, harm, alternative, decision owner, duration, reviewer and correction. It joins MOSAIC’s federated evidence to the powers in the proposed Act. It also preserves the proposition upon which this series has insisted from the beginning: law which cannot act before an irreversible attack is ineffective, but power which cannot say what it knows, why it believes it and what would make it stop is not defence. It is merely another unregulated force.

Bibliography

Books, Chapters and Articles

Aust HP, *Complicity and the Law of State Responsibility* (CUP 2011)

Avant DD, *The Market for Force: The Consequences of Privatizing Security* (CUP 2005)

Bethlehem D, 'Self-Defense against an Imminent or Actual Armed Attack by Nonstate Actors' (2012) 106 *American Journal of International Law* 770

Buchan R, 'Cyberspace, Non-State Actors and the Obligation to Prevent Transboundary Harm' (2016) 21 *Journal of Conflict and Security Law* 429

Buchanan B, *The Cybersecurity Dilemma: Hacking, Trust and Fear between Nations* (OUP 2017)

Coco A and de Souza Dias T, 'Cyber Due Diligence: A Patchwork of Protective Obligations in International Law' (2021) 32 *European Journal of International Law* 771

Corten O, *The Law Against War* (2nd edn, Hart 2021)

Crawford J, *State Responsibility: The General Part* (CUP 2013)

Dawidowicz M, *Third-Party Countermeasures in International Law* (CUP 2017)

Dinniss HH, *Cyber Warfare and the Laws of War* (CUP 2012)

Egloff FJ and Smeets M, 'Publicly Attributing Cyber Attacks: A Framework' (2023) 46 *Journal of Strategic Studies* 502 https://doi.org/10.1080/01402390.2021.1895117

Egloff FJ, 'Public Attribution of Cyber Intrusions' (2020) 6 *Journal of Cybersecurity* tyaa012 https://doi.org/10.1093/cybsec/tyaa012

Eichensehr KE, 'Decentralized Cyberattack Attribution' (2019) 113 *AJIL Unbound* 213 https://doi.org/10.1017/aju.2019.35

Eichensehr KE, 'The Law and Politics of Cyberattack Attribution' (2020) 67 *UCLA Law Review* 520

Finnemore M and Hollis DB, 'Beyond Naming and Shaming: Accusations and International Law in Cybersecurity' (2020) 31 European Journal of International Law 969

Green JA, 'Fluctuating Evidentiary Standards for Self-Defence in the International Court of Justice' (2009) 58 International and Comparative Law Quarterly 163

Greenwood C, 'International Law and the Pre-emptive Use of Force: Afghanistan, Al-Qaida, and Iraq' (2003) 4 *San Diego International Law Journal* 7

Hagemeyer-Witzleb TM, *The International Law of Economic Warfare* (Springer 2021)

Hathaway OA and others, 'The Law of Cyber-Attack' (2012) 100 *California Law Review* 817

Jensen ET and Watts SP, 'A Cyber Duty of Due Diligence: Gentle Civilizer or Crude Destabilizer?' (2017) 95 *Texas Law Review* 1555

Keitner CI, 'Attribution by Indictment' (2019) 113 *AJIL Unbound* 207 https://doi.org/10.1017/aju.2019.34

Kulesza J, *Due Diligence in International Law* (Brill Nijhoff 2016)

Mačák K, 'Decoding Article 8 of the International Law Commission's Articles on State Responsibility: Attribution of Cyber Operations by Non-State Actors' (2016) 21 Journal of Conflict and Security Law 405

Maurer T, *Cyber Mercenaries: The State, Hackers, and Power* (CUP 2018)

Rid T and Buchanan B, 'Attributing Cyber Attacks' (2015) 38 *Journal of Strategic Studies* 4 https://doi.org/10.1080/01402390.2014.977382

Roscini M, *Cyber Operations and the Use of Force in International Law* (OUP 2014)

Roscini M, 'Evidentiary Issues in International Disputes Related to State Responsibility for Cyber Operations' (2015) 50 Texas International Law Journal 233

Ruys T, *'Armed Attack' and Article 51 of the UN Charter: Evolutions in Customary Law and Practice* (CUP 2010)

Schmid E and Erceiş AÖ, 'The Attribution of Omissions: Due Diligence in Cyberspace and State Responsibility' (2023) 33 *Swiss Review of International and European Law* 577

Schmitt MN (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd edn, CUP 2017)

Schmitt MN, 'In Defense of Due Diligence in Cyberspace' (2015) 125 *Yale Law Journal Forum* 68

Shany Y and Schmitt MN, 'An International Attribution Mechanism for Hostile Cyber Operations' (2020) 96 International Law Studies 196

Tsagourias N, 'Cyber Attacks, Self-Defence and the Problem of Attribution' (2012) 17 *Journal of Conflict and Security Law* 229

Tsagourias N and Farrell M, 'Cyber Attribution: Technical and Legal Approaches and Challenges' (2020) 31 European Journal of International Law 941

Tsagourias N and Middleton F, 'Fact-Finding and Cyber Attribution' in Buchan R, Franchini D and Tsagourias N (eds), *The Changing Character of International Dispute Settlement: Challenges and Prospects* (CUP 2023) 439
https://doi.org/10.1017/9781009076296.023