SAND2020-6216C

Proceedings of the INMM 61stAnnual Meeting
July12-16, 2020, Baltimore, MD, USA

Facilitating International Cooperation through a Metrics Framework for Comparative Assessment of Nuclear Arms Control Verification Methods

Ben Bonin, Eva Uribe, Jason Reinhardt, Michael Hamel

*Sandia National Laboratories**

Abstract

Many technical challenges remain for verifying future nuclear arms control agreements, particularly those involving accounting and potential dismantlement of individual warheads. The state of international cooperation and technical exchange on addressing these challenges is fragmented, with disparate initiatives and research projects across government, academia, and international organizations. While these efforts have yielded many valuable contributions, the community of practice is notably lacking in consensus performance metrics for assessing the comparative advantages or disadvantages of alternative technologies and approaches. To draw a metaphor, there is no "miles per gallon" benchmark against which researchers can assess the degree to which any single technology, or set of technologies, is advancing progress toward broader verification goals and end states. Besides contributing to a lack of continuity across research agendas, this absence of metrics is also a barrier to achieving common understandings and objectives among international partners whose perspectives may be very different owing to unique political and cultural contexts; it is also a problem for policymakers as they weigh different verification options in arms control negotiations and set future R&D priorities. Drawing on previous metrics-relevant studies, this paper proposes core elements of a systems framework for deriving verification performance metrics that can serve as a basis for international dialog and technical exchange. An accompanying paper describes an in-depth approach to metrics development for information barriers using information theory and probabilistic analysis (Reinhardt et al 2020).

Introduction: Arms Control Challenges & Cooperation

Verification supported by monitoring technologies has been an essential component of nuclear arms control for nearly seventy years, beginning with emergence of the IAEA Safeguards system in the 1950s. While early U.S.-Soviet arms control treaties only allowed for monitoring by national technical means, the 1987 Intermediate Nuclear Forces (INF) Treaty introduced more intrusive onsite inspections supported by monitoring technologies. Since then, verification has been

* Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

foundational to U.S.-Russia arms control agreements including the Strategic Arms Reduction Treaty (START) and its successor New START. Even though the Comprehensive Test Ban Treaty (CTBT) has not entered into force, the International Monitoring System (IMS) developed to verify compliance with the agreement has played an important role in gathering data on nuclear tests carried out across the world, and in providing assurances that states parties remain faithful in their commitments to not test (Rueckert 1998, Tulliu 2003).

With the recent demise of the INF Treaty, and uncertainty regarding a successor to New START, there is reason for pessimism regarding the future of verified, treaty-based arms control. However, the history of arms control is not one of unbroken progress and continuity. The progress of negotiations and treaties has always ebbed and flowed, as evidenced by the emergence and demise of ambitious nuclear disarmament proposals in the late 1940s, including the Baruch Plan; the historic success of the first Strategic Arms Limitation Treaty (SALT I), followed by the failure of its successor, SALT II; and the Cold War-finishing triumph of INF and START, followed by the failure of START II and the eventual signing of New START.

Importantly, throughout these boom-and-bust periods, scientists and engineers worked to maintain dialog and collaboration on the technical dimensions of arms control, including verification and monitoring. The history of arms control on nuclear testing is notable for such collaboration, particularly the Group of Scientific Experts (GSE) formed in 1976 to specify the characteristics of an international monitoring system for a future test-ban treaty. In the two decades prior to CTBT opening for signature, the GSE carried out studies and tests that were ultimately foundational for the IMS (Dahlman 2020). This spirit of technical progress and collaboration on arms control challenges continues in the present, even where near-term prospects for a treaty appear remote. Governments, non-government organizations (NGOs), and academics continue to pursue ambitious research initiatives aimed at advancing new technologies and approaches.

There is no shortage of challenges for this work to address, especially if future agreements are to pursue potentially deeper cuts in nuclear arsenals – including non-strategic weapons – that may require accounting, inspection, and perhaps even dismantlement of individual warheads, rather than delivery systems. These include (in no particular order) establishing and confirming declaratory baselines; confirming the presence of accountable items, most notably warheads for which no consensus approach exists; maintaining the provenance of accountable items through transport, storage, and dismantlement; confirming the final disposition of accountable items; confirming that allowable deployed systems do not incorporate prohibited functionality; protecting sensitive information during inspections, both for national security and nonproliferation objectives; confirming that legitimate production operation do not create prohibited items or material; and detecting undeclared or clandestine items and activities (Cliff 2010, Fuller 2010, NRC 2005).

These challenges are not new, and some of them continue to grow in complexity with new technological advances that make strategically significant military capabilities accessible to a growing number of countries. Importantly, the solutions to these problems require sustained international dialog and cooperation, both because they are complex and because for any solution to be accepted as legitimate for use in verification of a bilateral or multilateral treaty, there must be a sense of international consensus.

The Need for Arms Control Verification Metrics

Applications of nuclear technology, both civilian and military, depend on often exquisite attention to precision measurement and clear performance metrics for safety, security, and reliability. Nuclear reactors are designed to meet performance thresholds in terms of energy output, reliability over a given operating period, and safety under certain adverse conditions. Nuclear weapons must meet strict performance standards for reliably operating when employed at the behest of national leadership, while never detonating under any other accidental or intentional circumstance. It is striking that in the field of nuclear arms control verification, similarly robust metrics are generally lacking. To draw a metaphor from the auto industry, there is no "miles per gallon" benchmark against which researchers from diverse international backgrounds can assess the degree to which any single technology, or set of technologies, is advancing progress toward broader verification goals and end states.

A metric is generally defined as a system of measurement that includes the item being measured, the unit of measurement, and the value of the unit. Metrics serve an important role in technology development. They provide a means by which to communicate progress, whether toward incremental gains, a specified target, or a desired end state. Metrics also provide a transparent and consistent means by which to compare different options and approaches. Importantly, metrics that are standardized, with agreement across international communities of practice, facilitate more effective communication and information sharing (Geisler 1999, NAS 2005). Returning to nuclear energy, the VVER-1000 and AP-1000 reactors may have been developed on opposite ends of the world by engineering communities speaking different languages, but the merits of the two designs can be readily debated using metrics accepted by nuclear science and engineering institutions across the world – thermal efficiency, megawatts output, and operating cost, to name a few.

The arms control field is not entirely devoid of metrics. The IAEA safeguards system employs multiple measurement concepts in evaluation of safeguards technologies and approaches. These include the ability to detect a "significant quantity" of diverted material equivalent to the minimum amount required for a nuclear explosive device. This detection must meet timeliness thresholds, determined by the minimum amount of time required to convert fissile material into weapon-useable form. Further metrics include probability of detecting diversion and probability of a false alarm. These metrics factor not only into evaluation of individual technologies and approaches, but also the overall safeguards implementation plan for a given member state. While there is contention around appropriate or ideal benchmarks for these metrics, the fact that they exist in the first place allows for a debate to take place under consistent terms of reference (Cochran 2007, IAEA 2001).

The safeguards system benefits from decades of experience, a reasonably well-resourced international institution employing a dedicated cadre of technical experts, and strong communities of practice across member states. Other initiatives supporting arms control verification, notably bilateral nuclear arms reduction treaties, have not enjoyed the same level of sustained attention and resources. The record suggests that technical verification advancements in support of U.S.-Soviet and U.S.-Russia arms reduction treaties were largely responsive to political developments; as treaties were negotiated, scientists and engineers were called upon to submit their best-available technical solutions, and in some cases improvise new solutions when needed. Fortunately, these treaties (which used counted delivery vehicles as a proxy for deployed warheads) did not require technologists to contend with many of the thornier challenges related to warhead verification and allowed for ready adaptation of off-the-shelf monitoring technologies.

Ambitious multilateral initiatives have periodically attempted to elevate and sustain attention toward nuclear warhead verification challenges including U.S.-Soviet and U.S.-Russia technical cooperation during the late 1980s and early 1990s under the auspices of various initiatives, the IAEA Trilateral Initiative from 1996-2002, the 2005-2015 U.K.-Norway Initiative, and most recently the International Panel on Nuclear Disarmament Verification and Quad Nuclear Verification Partnerships. While separated in time, these efforts have sought to leverage one another's results to the extent possible, often drawing from similar subject matter expert pools and addressing common technical challenges – albeit from sometimes different perspectives and at varying levels of technical depth (Hecker 2016, IPNDV 2017 & 2019, Plant 2019, Shea 2001).

These and other initiatives have yielded proposals and even testbed demonstrations for radiation measurement systems, tags and seals, managed access protocols, information barriers, and attribute templates, among other technologies and approaches. What is not clear, especially to an external observer, is how to judge the merits of competing approaches, or whether a new proposal truly represents an advancement on its predecessors. For example, when presented with two technical alternatives for an information barrier, how should performance be differentiated? What is the metric for information leakage that the two barriers are ostensibly intended to minimize? Is there a minimum threshold? And at a systems level, to what marginal extent does a higher performing barrier actually increase the confidence of the inspecting party when considered within a broader context of verification data collected to make a compliance determination?

Frameworks for Arms Control Verification Metrics

This need for metrics has been recognized and addressed to varying degrees by previous studies and initiatives, many of which have been funded by the U.S. government, and a smaller number of which have emerged from international partnerships. While these efforts tackled the metrics problem from differing perspectives and to varying degrees of technical depth, there are notable common methodological threads.

A 2014 Defense Science Board (DSB) task force report addressing the role of technology in supporting future arms control and nonproliferation initiatives outlined an approach to assessment of alternatives using scenario-based functional decomposition. The approach begins with a detailed scenario mapping (see Figure 1), identifying adversary pathways through key process “nodes” leading to potential scenario end states (e.g. an adversary defecting from its treaty obligations). These nodes represent potential opportunities along the scenario pathway to implement monitoring solutions for detection of aberrant behavior. For a given node, key strategic capability areas can be identified and further decomposed into objectives, tasks, and ultimately assets (i.e. monitoring protocols, personnel, and technology). Performance metrics can be identified at each level of this decomposition, ideally allowing for asset-level metrics to be “rolled up” into higher level metrics supporting overall evaluation of the monitoring system. In their call for a paradigm shift, the taskforce acknowledged that metrics for determining successful monitoring technologies depend on the context in which they are to be used, including the overarching objectives (e.g., verifying warhead dismantlement vs. detecting a nascent nuclear weapons program) as well as the parties involved (DSB 2014).

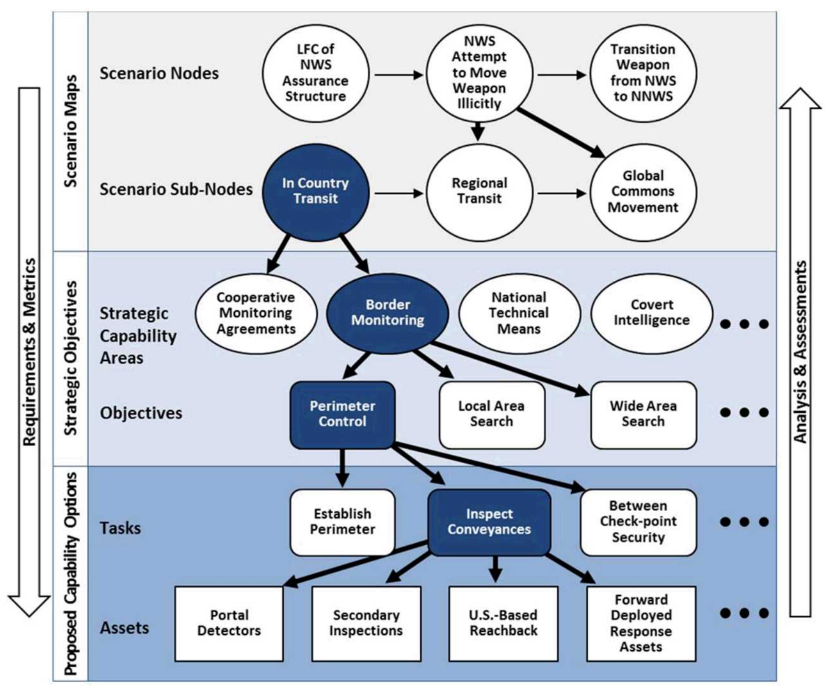

Figure 1: Example scenario decomposition map from DSB report *Assessment of Nuclear Monitoring and Verification Technologies* (DSB 2014, pp 45)

A 2016 report detailing a collaborative effort between Los Alamos National Laboratory, Lawrence Livermore National Laboratory, Sandia National Laboratories, and Pacific Northwest National Laboratories outlines a system evaluation methodology for assessing the effectiveness of proposed monitoring approaches for future stockpile reduction treaties. The effort focused on three basic verification objectives under a hypothetical agreement including confirming a declared number of treaty accountable items, confirming that the treaty accountable items are warheads, and confirming the dismantlement of those warheads. The evaluation framework included a decomposition process that began with establishing high level monitoring objectives, identifying a hierarchical set of functions supporting those objectives, and connecting those functions to technology options that could be evaluated quantitatively across a set of performance and effectiveness metrics and for a range of verification scenarios. Example metrics identified through this framework included time-to-detection of a discrepancy and time-to-confidence in determining the magnitude of the discrepancy, in addition to technology-level metrics like false positive and false negative rates (Chen et al. 2016).

The International Panel on Nuclear Disarmament Verification (IPNDV) is an ongoing multilateral effort involving more than 25 countries, dedicated to identifying and overcoming the principle challenges of nuclear disarmament verification. A Working Group on Monitoring and Verification Objectives has identified core principles that should underly the weapons dismantlement process – some of which lend themselves to metrics development – including effectiveness, building confidence, nonproliferation, non-interference, cost-efficiency, determinacy, and structure. In addition, the IPNDV as outlined a 14-step process framework (see Figure 2) illustrating the most likely stepwise progression of a generic disarmament process. For certain steps in the dismantlement phase, the group has undertaken a thorough review and gap analysis of existing technologies and approaches, including a subject matter expert-informed assessment of strengths and limitations (IPNDV 2017 & 2019).

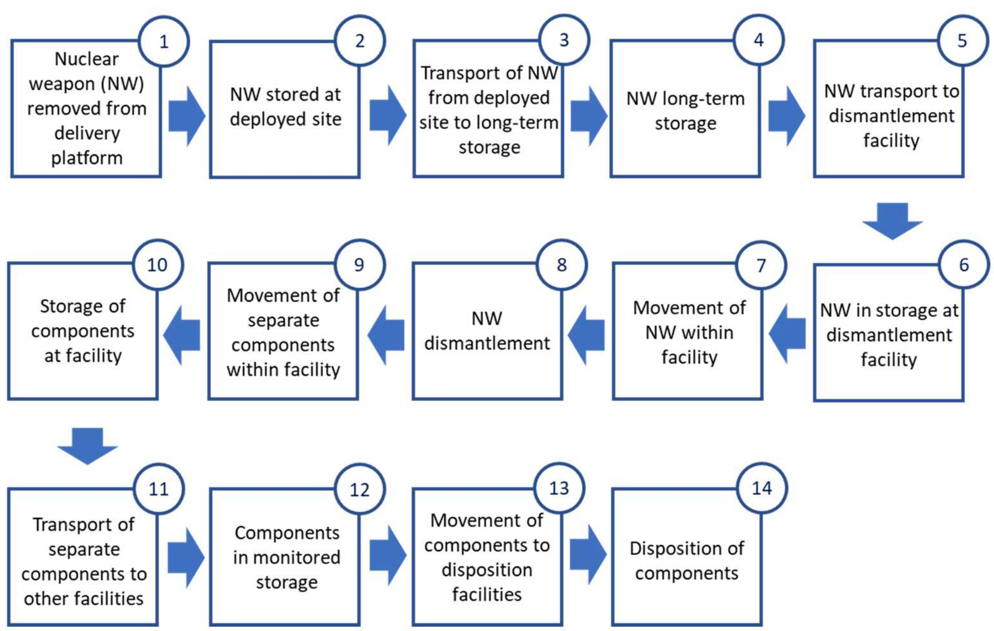

Figure 2: IPNDV 14-step verification framework (adapted from IPNDV 2017, pp 11)

The efforts described in this section represent some of the most recent and public-facing examples of metrics-relevant work conducted in the verification field; no doubt they only represent a subset of thinking that has taken place on the topic. However, at least three common themes are evident even from this modest sampling. The first is that metrics for comparing monitoring and verification approaches should derive from a core set of fundamental, high-level objectives. This is exemplified by the seven principles described by the IPNDV, and the three monitoring and verification objectives highlighted by the multilab system evaluation team. The second theme is that these high-level objectives should be connected to specific technologies or processes through a functional hierarchy or decomposition. The multilab effort provided a notional example of such a functional decomposition. Similarly, the DSB taskforce report contains examples of scenario decomposition to derive objectives, tasks, and assets or measurement approaches that are specific to a scenario. The third theme is that functional decomposition and/or metrics should be tailored to specific scenarios, that is, to specific stages in the dismantlement process and to specific technology or process "end-uses" as defined by the international partners involved. Scenario selection is integral to all the methodologies reviewed; both the DSB and multilab efforts emphasize the importance of developing metrics in reference to a specific verification scenario, while the IPNDV 14-step framework makes explicit the scenario space under consideration.

Conclusion: A Proposal for International Dialog and Cooperation on Verification Metrics

The authors of this paper contend that international dialog and collaboration on verification technologies and approaches would benefit from a systematic, principled, and repeatable framework for deriving metrics that can ultimately be adopted across communities of practice. This framework should inform the development of verification performance metrics that, to the greatest extent possible, are technology- and protocol-agnostic, facilitating their application across a range of technology solutions and potential agreements. Importantly, this effort does not have to start from scratch. The work highlighted in the previous section provides ample precedent and basic organizing concepts, if not necessarily a template that can be adopted whole cloth.

For purposes of international dialog and cooperation, a workable metrics framework will likely need to meet several criteria. It will need to be conceptually intuitive for purposes of communicating across language barriers and adapting to different cultural contexts. It should be scalable and adaptable across a range of technical contexts likely to be encountered in any number of plausible future arms control agreements. It should consider the needs of both technical communities of practice and the non-technical decisionmakers who might use these metrics to decide between different verification and monitoring options as treaty protocols are negotiated.

Most importantly, the basic parameters of such a framework should be arrived at by international consensus, rather than presented as a fait accompli by even well-intentioned facilitators. In that spirit, such a dialog might be structured around the following tasks, informed by previous research and initiatives:

> **Define the scenario space.** Given the potential diversity of verification scenarios that might be encountered under future agreements, and the potential context-dependency of certain metrics, international partners might begin with identifying a standardized approach

for defining and bounding the verification scenario space. Both the IPNDV 14-step framework as well as the DSB taskforce scenario decomposition process provide starting points for adaptation or modification.

Identify high-level monitoring and verification objectives. To the extent possible, partners should strive to identify a core set of fundamental verification objectives likely to be relevant across a range of plausible arms control agreements, though scenario-specific objectives will likely also be needed. Inspiration for these objectives can be found in the IPNDV verification principles.

Map verification technology and approaches to objectives. A consensus, systematic approach to drawing connections between verification technologies/approaches and objectives will facilitate consistency and repeatability in how metrics are derived; it will also facilitate potential aggregation of metrics allowing for assessment of multi-technology verification approaches and regimes, as opposed to just singular technologies. The most logical point of departure for this appears to be a functional decomposition process as illustrated by the DSB task force and the multilab systems evaluation study.

Develop and evaluate metrics. With a functional decomposition in place, metrics can begin to be identified at various levels of the decomposition, evaluating progress toward objectives at the highest level, and technical performance benchmarks at the lowest. The multilab systems evaluation methodology, along with the experience of IAEA safeguards, provides examples of what such metrics might look like.

Converge on a research agenda. By conducting this process several times across a range of relevant scenarios, participants will begin to converge upon a research agenda that is guided by a common understanding of the key principles and challenges and can be evaluated and compared across a common set of criteria.

An international conversation on metrics can potentially take place in any number of forums. The topic is complicated enough, with implications across the entire breadth of existing arms control initiatives, that it could be the basis for a dedicated dialog. At the same time, it might also be broached in the context of existing forums like the IAEA or the IPNDV. The authors of this paper are agnostic in this regard, deferring to governments, international organizations, and communities of practice on the most appropriate context for a dialog.

Arms control will always be subject to uncertainty in the international political environment. However, research and exchanges on verification technical challenges provide an opportunity to continue advancing the science and engineering of arms control, laying the groundwork for more robust verification regimes with strong international buy-in. Moreover, history demonstrates that sustained outreach and community building can grow confidence and mutual trust between nations that is necessary both to catalyze progress on existing agreements and to spark future negotiations. As an effort toward strengthening bonds between arms control communities of practice, reaching a consensus on metrics would represent a modest but significant contribution toward greater continuity and coordination in the pursuit of diverse verification research agendas.

References

Chen, C. et al. *Developing a System Evaluation Methodology for a Warhead Monitoring System*. Sandia National Laboratories, Report SAND2016-9371C. 2016.

Cliff, D., H. Elbahtimy, & A. Persbo. "Verifying Warhead Dismantlement: Past, Present, Future." Verification Research, Training and Information Centre (VERTIC). 2010.

Cochran, C. "Adequacy of IAEA's Safeguards for Achieving Timely Detection." In *Falling Behind: International Scrutiny of the Peaceful Atom*, edited by H. Sokolski. Washington, DC: Nonproliferation Policy Education Center. 2007.

Dahlman, O. et al. "The inside story of the Group of Scientific Experts and its key role in developing the CTBT verification regime." *The Nonproliferation Review* 27, no. 1/2. 2020.

Department of Defense Science Board (DSB). *Task Force Report: Assessment of Nuclear Monitoring and Verification Technologies.* Washington, DC: Defense Science Board. 2014.

Fuller, J. "Verification on the Road to Zero: Issues for Nuclear Warhead Dismantlement." *Arms Control Today* (December). pp 19-27. 2010

Geisler, E. "The Metrics of Technology Evaluation: Where We Stand and Where We Should Go from Here." Presented at the 24th Annual Technology Transfer Society Meeting, July 15-17, 1999. pp 5-7. 1999.

Hecker, S. *Doomed to Cooperate: How American and Russian Scientists Joined Forces to Avert Some of the Greatest Post-Cold War Nuclear Dangers*. Los Alamos: Bathtub Row Press. 2016.

International Atomic Energy Agency (IAEA). *IAEA Safeguards Glossary*. International Nuclear Verification Series No. 3. 2001.

International Partnership for Disarmament Verification (IPNDV). *Phase I Summary Report: Creating the Building Blocks for Future Nuclear Disarmamen*t. Washington, DC: Nuclear Threat Initiative. 2017.

International Partnership for Disarmament Verification (IPNDV). *Phase II Summary Report: Moving from Paper to Practice in Nuclear Disarmament Verification.* Washington, DC: Nuclear Threat Initiative. 2019.

National Academy of Sciences (NAS). *Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program*. Washington, DC: The National Academies Press. pp 14-16. 2005.

National Research Council (NRC). *Monitoring Nuclear Weapons and Nuclear Explosive Materials: An Assessment of Methods and Capabilities*. Washington, DC: The National Academies Press. 2005.

Plant, Tom. "The Disarmament Laboratory: Substance and Performance in UK Nuclear Disarmament Verification Research." Finnish Institute of International Affairs, Working Paper 111. 2019.

Reinhardt et al. "Information Theoretic Measure of Verification System Performance." Proceedings of the INMM 61st Annual Meeting, July12-16, 2020, Baltimore, MD. 2020.

Rueckert, G. *On-site Inspection in Theory and Practice: A Primer on Modern Arms Control Regimes*. Praegor: Westport, Connecticut. 1998.

Shea, T. *Report on the Trilateral Initiative: IAEA Verification of Weapon-origin Material in the Russian Federation and the United States*. International Atomic Energy Agency Bulletin. 2001.

Tulliu, S. and T. Schmalberger. *Coming to Terms with Security: A Lexicon for Arms Control, Disarmament, and Confidence-building*. Geneva: United Nations Institute for Disarmament Research. pp 225. 2003.