

Princeton Plasma Physics Laboratory

PPPL-

PPPL-



Prepared for the U.S. Department of Energy under Contract DE-AC02-09CH11466.

Princeton Plasma Physics Laboratory Report Disclaimers

Full Legal Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Trademark Disclaimer

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors.

PPPL Report Availability

Princeton Plasma Physics Laboratory:

<http://www.pppl.gov/techreports.cfm>

Office of Scientific and Technical Information (OSTI):

<http://www.osti.gov/bridge>

Related Links:

[U.S. Department of Energy](#)

[Office of Scientific and Technical Information](#)

[Fusion Links](#)

A Zero-Knowledge Protocol for Nuclear Warhead Verification

Alexander Glaser,* Boaz Barak,** Robert J. Goldston***

*Program on Science and Global Security, 221 Nassau St, Fl 2, Princeton University, Princeton, NJ 08542; **Microsoft Research, 1 Memorial Drive, Cambridge, MA 02142; ***Princeton Plasma Physics Laboratory, Princeton University, P.O. Box 451, Princeton, NJ 08540

Summary: The verification of nuclear warheads for arms control faces a paradox: international inspectors must gain high confidence in the authenticity of submitted items while learning nothing about them. Conventional inspection systems featuring “information barriers,” designed to hide measurements stored in electronic systems, are at risk of tampering and snooping. Here we show the viability of a fundamentally new approach to nuclear warhead verification that incorporates a zero-knowledge protocol, designed such that sensitive information is *never measured* so does not need to be hidden. We interrogate submitted items with energetic neutrons, making, in effect, differential measurements of neutron transmission and emission. Calculations of diversion scenarios show that a high degree of discrimination can be achieved while revealing zero information. Timely demonstration of the viability of such an approach could be critical for the next round of arms-control negotiations, which will likely require verification of individual warheads, rather than whole delivery systems.

1. Introduction

Existing nuclear arms-control agreements between the United States and Russia place limits on the number of *deployed* strategic nuclear weapons. Verification of these agreements take advantage of the fact that deployed weapons are associated with unique and easily accountable delivery platforms, i.e., missile silos, submarines, and strategic bombers, to which agreed numbers of warheads are attributed. The next round of nuclear arms-control agreements, however, may place limits on the

total number of nuclear weapons and warheads in the arsenals.^a This would include tactical weapons as well as deployed and non-deployed weapons. Such agreements would require new verification approaches, including inspections of individual nuclear warheads in storage and warheads entering the dismantlement queue. This is a qualitatively new challenge because the design of nuclear weapons is highly classified information that cannot be exposed to international inspectors. A viable verification approach therefore has to resolve the tension between reliably verifying that the inspected warhead is *authentic* while avoiding disclosure of information about its design.^{1,2,3,4}

Practitioners and policy makers have been well aware of this conundrum, and prior work by national laboratories in the United States, Russia, and the United Kingdom addressed it by using “information barriers.”^{2,4} These barriers consist of sophisticated automated systems that process highly classified information measured during an inspection, but only display results in a yes/no manner. Such systems are inherently complex, and require both parties to trust that they have no trapdoors hidden from the inspector, which could be used to cause a system to declare invalid objects as authentic, nor side channels unknown to the host, which could leak classified information to the inspector or others. These concerns are serious obstacles to adopting such systems.

In this work we consider a fundamentally different approach to this problem. Rather than trying to acquire and analyze classified data behind an engineered information barrier, we use the cryptographic notion of *zero knowledge proofs* to ensure that sensitive data is never measured in the first place.

^a Upon signing the New START Treaty in April 2010, U.S. President Obama said: “While the New START treaty is an important step forward, [...] we hope to pursue discussions with Russia on reducing both our strategic and tactical weapons, including non-deployed weapons.”

2. Zero-knowledge Proofs (with Marbles)

Zero-knowledge proofs, invented in the 1980's by Goldwasser, Micali, and Rackoff,⁵ have become an important tool of modern cryptography. Zero-knowledge proofs achieve the paradoxical goal of allowing one to prove that a statement is true without revealing *why* it is true. Such proofs are extremely useful for many digital applications, including privacy-preserving data mining, electronic voting, and online auctions.⁶ To achieve zero knowledge, Goldwasser et al. extended the traditional notion of a proof from a static text to a *protocol*, which involves *randomization* and *interaction* between the prover and verifier. At the end of the protocol, the verifier has a high degree of confidence that the statement is correct, while the prover is guaranteed that the verifier did not learn anything about the data underlying the truth of the statement. For our application, the host submitting warheads for inspection takes the role of the prover and the inspector the role of the verifier.

While classical zero-knowledge proofs are *digital* protocols, proving statements about mathematical objects, we illustrate the concept using a *physical* zero-knowledge protocol that is closely related to our proposed verification approach (Figure 1):

Alice (the host) has two small cups both containing X marbles where X is some number between 1 and 100. She wants to prove to Bob (the inspector) that both cups contain the same number of marbles, without revealing to him what this number X is. To do so, Alice prepares two buckets, which she claims each contain $(100 - X)$ marbles. Bob now randomly chooses into which bucket which cup is poured. Once this is done, Bob verifies that both buckets contain 100 marbles.

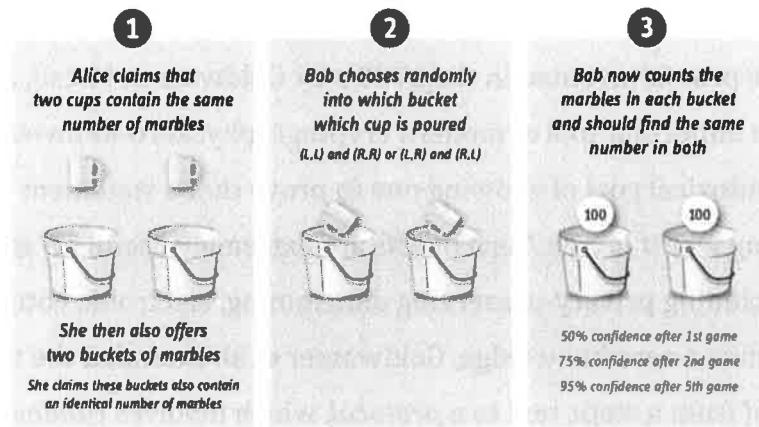


Figure 1. A zero-knowledge protocol to prove that two cups contain the same number of marbles. The confidence level increases with the number of games played.

The protocol reveals no information on X since, regardless of X's value, Bob always sees 100 marbles in both buckets. However, if the cups did not have the same number of marbles, then no matter how Alice prepares the buckets, with probability of 50% after the pouring one of the buckets will not contain 100 marbles. If Alice and Bob repeat this game, say, five times, then if Alice consistently cheats she will be caught with $(1-2^{-5}) > 95\%$ probability.

3. From Marbles to Neutrons

The analogy to our setting is that we want to show that two or more putative warheads have identical neutron transmission and emission counts under irradiation by high-energy neutrons. We follow the *template approach* for warhead verification,² where a radiation measurement generates a complex and unique fingerprint of an inspected item.^b This fingerprint is then compared against the

^b The other approach for verification is known as the *attribute approach*, where the system checks that the inspected item satisfies certain agreed upon properties such as containing at least X grams of plutonium. While the attribute approach has the advantage of not needing access to a classified template, it is inherently limited. For example, since X has to be an unclassified lower bound, the true amount of plutonium would probably be significantly higher, and such a system will not flag removal

fingerprint of one or more templates to confirm that all items are materially identical. Templates can be directly selected from deployed weapons so that the inspector has high confidence in their authenticity. In the case of weapon systems that are not currently deployed, one could gain confidence by measuring a large number of warheads that were known by national technical or other means to have been in storage for a long time.⁷

We compare the submitted items by recording the transmission pattern of 14-MeV neutrons, as well as recording the intensity of neutrons emitted to the sides of the items. Neutron radiographic images of warheads contain highly classified information, but in our case they are actually never measured. Rather, in analogy to the marbles example, they are recorded using detectors that are *preloaded* with the negative of the radiograph. Preloaded values are not revealed to the inspector. As in the marbles example, after the measurement and if the host is telling the truth, the inspector always sees the same number of counts in every detector. Furthermore, as in the marbles example, preloads supplied with the submitted items are shuffled at random, so if the items actually differ, and the preloads are chosen to complement the differing items, then with significant probability the image will *not* be uniform, and a mismatch will be present on both items.

Unlike the marbles example, neutron measurements are inherently statistically noisy. To avoid conveying information through the noise distribution we use preloaded values that are noisy as well. In particular, since the measurement distribution will be Poisson, we use a Poisson distribution for the preloads, and, using the fact that the sum of two Poisson distribution is also Poisson, our protocol achieves the following:

of any quantity in excess of the agreed minimum. Note that all prior systems (both attribute and template based) have required an engineered information barrier.

The neutron count obtained by any measurement on the template or on a valid submitted item is distributed according to the Poisson distribution with mean and variance equal to a previously agreed upon value N_{max} .

Since N_{max} is known in advance to both sides, neither the measurement nor its noise reveals any new information. N_{max} for transmission could reasonably correspond to the maximum number of counts that is expected in the absence of a test item. If a submitted item varies from the true warhead (or the submitted preloads are not identical) an image may be seen that could contain sensitive information. This will be an additional strong incentive for the host not to cheat.

For simplicity of operation, we envision that the host places the detectors for each measurement in a removable board that forms part of the measurement system. Crucially, the inspector chooses which board to use with which test item. As in the marbles example, this means that if the host uses unmatched boards to try to mask invalid items, then with 50% probability the invalidity will be made *more evident* by the measurement with the mismatched boards. Since we expect that this “game” will be repeated many times, even a risk-tolerant host would not accept the resulting low chance of success. We note that testing multiple warheads in parallel is an attractive option, because it makes the probability of detecting the use of non-identical preloads significantly higher.

Once the measurements have been completed and the detectors read out, the inspecting party can verify the functionality of the detectors by exposing them to additional neutrons. This is an important advantage of the proposed method. A pioneer in this field, James Fuller, recently stated that *“after all these years, no one has yet demonstrated either an attribute or template type system using a classified test object in such a way that specialists from the inspecting country can then [i.e., after a measurement] thoroughly examine and proof the measurement equipment.”*⁸

While we examine here neutron measurements using preloaded non-electronic detectors, there may be other non-electronic zero-knowledge protocols for warhead verification that can avoid the use of engineered information barriers. Indeed such systems could be complementary to the neutron measurements discussed here.

4. Monte Carlo Analysis

We now show how our approach can be implemented in practice, and that “diversions” between objects can be reliably detected. We have analyzed the approach with a series of MCNP5 Monte Carlo simulations.⁹ Construction of a physical experimental setup is underway.

We propose to use 14-MeV neutrons from a DT neutron generator¹⁰ to interrogate test items, allowing detailed transmission profile measurements and also measurements of neutron intensities at large angles due to elastic and inelastic scattering, fission, and (n,2n) reactions. The neutrons from the generator are collimated by 60 cm of polyethylene and illuminate the inspected item (Figure 2). An array of neutron detectors placed at a distance of 50 cm behind the center of the item provides the transmission measurements. Additional detectors (not shown) can be positioned at large angles to the beam, i.e., in the shadow of the collimator, to measure neutrons emitted from the test item.

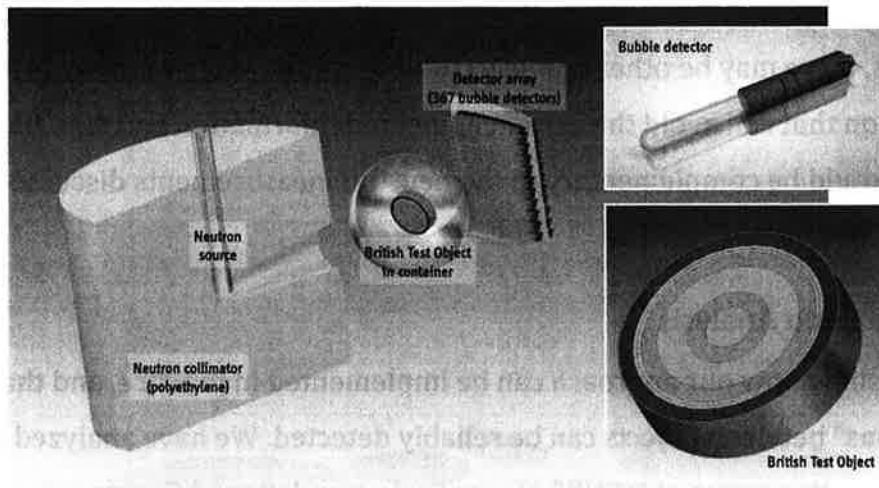


Figure 2. Experimental setup with neutron source, test item in container, and detector array (left). Insets show typical bubble detectors (top) and the British Test Object (bottom). Large-angle detectors are not shown. *3D models: Sébastien Philippe, Princeton University; Bubble detector design: Bubble Technology Industries.*

Test item. The test item used for this analysis is the unclassified “British Test Object” (BTO), which consists of concentric rings of different materials, including polystyrene, tungsten (two rings with a combined mass of 7.74 kg), aluminum, graphite, and steel. The BTO has an outer diameter of 18.9 cm and a height of 5 cm. This test object does not contain special or other nuclear materials, but is used to develop and calibrate imaging systems for diagnostic analysis of nuclear weapons.¹¹ The BTO is emplaced in a container in order to avoid revealing to the inspector the appearance or orientation of the inspected item inside the container.

Detector array. To assess the viability of our proposed protocol, we work with a board holding a hexagonal array of 367 detectors consisting of 21 rows of 17 or 18 detectors comprising an area of about $42 \times 42 \text{ cm}^2$. The assumed area of each detector (pixel) is 2 cm^2 . By rotating the BTO, the board can image it in any orientation.

In the analysis below, detectors are assumed to be sensitive to neutron energies > 10 MeV. Neutrons scattered from the walls of the room (“room return”) are not included in the calculations but their effect should be small given the energy cutoff, particularly if the room is specially prepared for the inspection, e.g. with borated polyethylene in front of borated concrete walls.

Figure 3 illustrates typical results on a template and a valid item in two different orientations; for reference purposes, the respective neutron radiographs of the test items are shown, but this data is *never measured* in the inspection, since only preloaded detectors are used. As expected, in the case of inspecting a valid item, detector counts are distributed consistent with a Poisson distribution with mean and variance N_{max} . In the following, only the more challenging side-view orientation of the BTO is used for an analysis of four representative diversion scenarios.

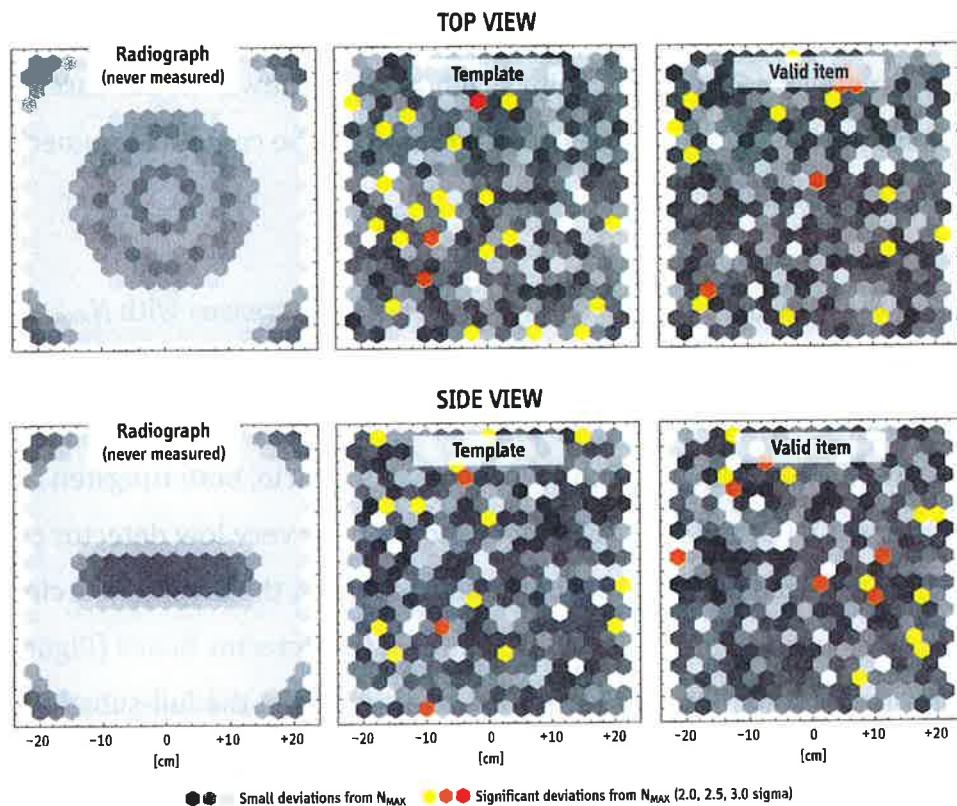


Figure 3. Results of MCNP5 simulations for interrogations of the British Test Object in two different orientations. Inspection of valid items reveals no information about them. The radiographs of the items

shown on the left are never measured; they correspond to measurements without preloading the detectors. The other panels visualize total detector counts, including the host's preloads, after the measurements on the template and on an identical (valid) item, and hence are simply independent Poisson random variables with expectation N_{max} . Shades of gray and colors indicate absolute differences from N_{max} . For this calculation, $N_{max} = 5,000$, but results would look the same for any choice of N_{max} .

Diversion scenarios. To examine diversion scenarios, we need to define a decision rule to distinguish passed from failed tests. For our present purposes we use a very simple rule looking for statistical outliers on predefined groups of pixels. If we denote individual detector counts by the numbers X_1, \dots, X_n , then we can define new numbers Y_1, \dots, Y_k , where every Y_j is the sum of a small number of the X_i 's, divided by the expected standard deviation of Y for a match case. We define the test to be *positive* (i.e., diversion detected) if there is at least one j with $|Y_j| > T$, where T is a threshold chosen such that in the match case for every j the probability that $|Y_j| > T$ is at most p_{fp}/k where p_{fp} is our allowed false positive rate. Concretely, in our setting, we examine $k = 295$ non-disjoint seven-pixel windows.^c In this case, to achieve a false positive rate $p_{fp} \leq 0.05$, the threshold can be computed numerically to be $T = 3.76$ standard deviations.

Sensitivity of the measurements to diversion scenarios increases with N_{max} and the associated improvements of counting statistics. We therefore examine a series of different diversion scenarios and a range of values for N_{max} below to determine system requirements (Table 1). In the full-removal scenario, both tungsten rings are removed from the BTO, which is easily detected even for very low detector counts. Similarly, if lead is used to substitute both tungsten rings, the diversion is clearly distinguishable even by simple visual inspection of the detector board (Figure 4, top). Our proposed statistical test identifies the diversion in the full-substitution

^c The detector bank has 367 positions but only 295 detectors have all six nearest neighbors. A seven-pixel window is defined here by a central pixel and its nearest neighbors.

scenario with a probability of true positives, $p_{tp} > 0.99$, even for N_{max} as low as 1,000 detector counts.

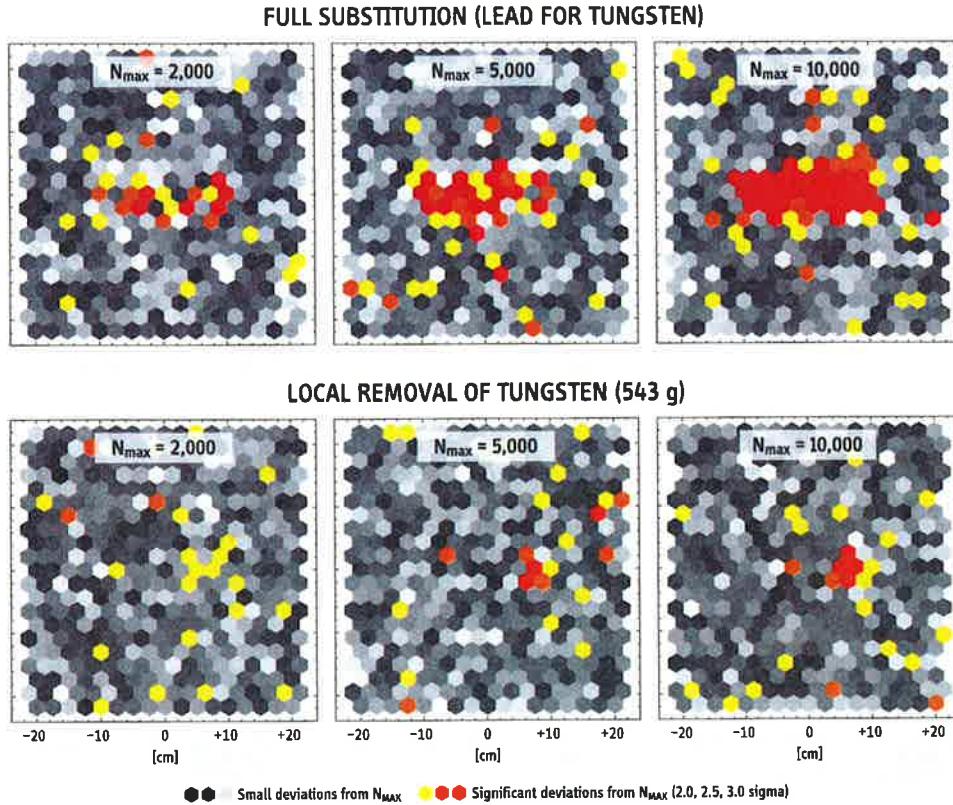


Figure 4. Results of MCNP5 simulations for two notional diversion scenarios. Inspection of invalid items results in anomalies of the detector counts that become more pronounced with increasing N_{max} . Shades of gray and colors indicate absolute differences from the selected values of $N_{max} = 2,000, 5,000$, and $10,000$.

The local-removal and local-substitution scenarios are more challenging. In these cases, a 36-degree sector of the outer tungsten ring is removed or replaced, which corresponds to a diversion of 543 g of tungsten contained in the BTO. To achieve a detection probability of 95%, an N_{max} of 5,000 is required in the case of the localized tungsten removal. When lead is used to substitute for tungsten in the 36-degree sector, N_{max} increases to 32,000 for the same detection probability.

Note that in these studies no use has been made of the side detectors. The more realistic case of substitution of ^{238}U for ^{235}U in a nuclear weapon component results

in a factor of about two reduction in the induced fission rate due to 14 MeV neutrons. Substitution of reactor-grade for weapon-grade plutonium has a small effect on the directly induced fission rate, but a large effect on the spontaneous fission rate, which could be detected by operating the side detectors in the absence of the neutron source. Thus the calculations presented here are conservative.

We note that 5% of the items will be flagged as invalid by our proposed test procedure due to the set 5% false positive rate, even in the case where all items are valid. Retesting flagged items will rapidly determine their validity. If a 5% false negative rate is deemed unacceptable, either routine retesting or a greater N_{max} can be implemented to reduce this value. The optimization of any retesting scenario, and study of a wider range of host strategies for cheating, as well as inspector strategies for analyzing signal patterns to find such cheating, will be the subject of future research.

N_{max}	500	1,000	2,000	5,000	10,000	32,000
Valid Item	$\leq 5\%$ (by design, in all cases)					
Full Removal	$> 99.9\%$ (in all cases)					
Full Substitution	77.7%	99.5%	> 99.9%	> 99.9%	> 99.9%	> 99.9%
Local Removal	<i>undetectable</i>	15.7%	41.7%	94.6%	> 99.9%	> 99.9%
Local Substitution	<i>undetectable</i>	<i>undetectable</i>	6.0%	11.7%	30.2%	95.5%

Table 1. Probabilities for an item to be flagged as “invalid” as a function of total expected count N_{max} .

5. Preloadable non-electronic detectors

Perhaps the most critical aspect of a viable implementation of the proposed verification approach is the choice of the detector technology. The detectors must have the capability to be preloaded with a desired neutron count prior to the inspection. At a minimum, this preload has to persist for hours or days and its decay rate, if present, be well characterized. Preloaded counts must be indistinguishable from counts accumulated during irradiation of the test items. Ideally, detectors should be energy selective so that the effect of low energy neutrons returning from room walls can be controlled most easily, be insensitive to γ 's, have high efficiency, and permit total counts in the range discussed above. Finally, relying on a non-electronic detection mechanism is highly advantageous given that complex electronic components and circuits are potentially vulnerable to tampering and snooping. We find that at least two detector technologies can meet these criteria: superheated emulsions ("bubble detectors") and neutron activation analysis detectors.

In superheated emulsions, neutron recoil particles trigger the formation of macroscopically observable bubbles from microscopic droplets that are dispersed in an inert matrix.¹² These detectors can be configured to have essentially any desired energy threshold between tens of keV and tens of MeV. Commercially available, polymer-based "bubble detectors" are limited to a maximum bubble count on the order of a few hundred bubbles, beyond which camera-based imaging techniques cannot resolve bubbles individually. "Superheated drop detectors" produced with a more compliant aqueous gel can be used up to much higher bubble counts. Either optical tomography or magnetic resonance imaging allow counting bubbles hidden in the depth of the fluid.^{13,14} If the highest N_{max} is desired, multiple detectors can be exposed in series. By the proper choice of compliant matrix, detectable aging (growth) of bubbles can be eliminated. Net detection efficiency in the order of 1% can be easily achieved. The emulsions can be contained in opaque containers so that a preload is not visible to the inspector.

For neutron activation analysis (NAA), imaging can be undertaken, for example, using an array of hexagonal prisms made of zirconium.¹⁵ ⁹⁰Zr has a neutron activation threshold of 12 MeV through an (n,2n) reaction. The resulting ⁸⁹Zr has a half-life of 3.3 days, which must be taken into account to determine the required level of preloading. Counting the γ rays from ⁸⁹Zr decay in high-purity germanium well detectors should give a net detection efficiency of about 0.25%. For 3 cm long prisms, with cross-sectional area of 2 cm², this would provide an N_{max} in the range of 20,000 within one hour, for a commercially available DT neutron generator producing 3×10^8 n/sec.¹⁶ Indium has an appropriate activation response to fission neutrons, with reduced sensitivity in the range of 14 MeV, for use in the side detectors. It will be important to assure that unshielded, preloaded activation samples are not in the presence of γ detectors before their final exposure.

Detectors can be preloaded with counts with the appropriate statistical properties by exposing them to energetic neutrons for a pre-calculated period of time and/or through a pre-calculated depth of shielding. As discussed above, statistical noise in the measurement will not reveal any information. However any systematic measurement errors must be well understood, such that while one detector may be characterized by a different efficiency than another, which can be calibrated out, this efficiency must not vary significantly between the preload and the measurement processes. For example, it is important to maintain control over the temperature of bubble detectors during irradiation. The DT neutron generator must also be well controlled and measurable, so that there is no significant variation in the shape of the neutron field produced nor in the total number of neutrons emitted when irradiating items. An accurate neutron flux monitor can be used to set the irradiation time, so perfect reproducibility is not required in the rate of neutron production. We anticipate that these requirements can be met, but the techniques to achieve the necessary degree of control need to be demonstrated and validated.

6. Conclusion

Authenticating nuclear warheads without revealing classified information represents a qualitatively new challenge for international arms-control inspection. Here we have shown an example of a zero-knowledge protocol based on non-electronic differential measurements of transmitted and emitted neutrons that can detect small diversions of heavy metal from a representative test object. This technique will reveal no information about the composition or design of nuclear weapons when only true warheads are submitted for authentication, and so does not require an engineered information barrier. Other such zero-knowledge protocols may be possible. The zero-knowledge approach has the potential to remove a major technical obstacle to new nuclear arms control agreements that include both deployed and non-deployed, strategic and tactical weapons, at substantially lower levels of armament than current agreements.

References

1. *Confidence, security & verification: The challenge of global nuclear weapons arms Control.* Atomic Weapons Establishment, Aldermaston, Crown, 2000.
2. Spears, D. (ed.) *Technology R&D for arms control.* U.S. Department of Energy, Office of Nonproliferation Research and Engineering, Washington, DC, Spring 2001, www.ipfmlibrary.org/doe01b.pdf.
3. Fuller J. Going to zero: Verifying nuclear warhead dismantlement. Chapter 4 in Hinderstein, C. (ed.) *Cultivating confidence: Verification, monitoring, and enforcement for a world free of nuclear weapons.* Hoover Institution Press, 2010.

4. Anderson, B., Beach, H., Finney, J., Ritchie, N., Saakyan, R. & Watson, C. *Verification of nuclear weapon dismantlement: Peer review of the UK MoD programme*. British Pugwash Group, London, November 2012.
5. Goldwasser, S., Micali, S. & Rackoff, C. The knowledge complexity of interactive proof-systems. *SIAM Journal on Computing* **18**, 186–200 (1989).
6. Chazelle, B. The security of knowing nothing, *Nature* **446**, 992–993 (2007).
7. Fuller J. Verification on the road to zero: Issues for nuclear warhead dismantlement. *Arms Control Today*, December 2010.
8. Fuller, J., personal communication, October 2012.
9. *A general Monte Carlo N-particle (MCNP) transport code*. Los Alamos National Laboratory, mcnp.lanl.gov.
10. Roquemore, A. L., Jassby, D. L., Johnson, L. C., Strachan, J.D. & Barnes, C. W. Performance of a 14-MeV neutron generator as an in situ calibration source for TFTR. *15th IEEE/NPSS Symposium on Fusion Engineering*, 114–118 (1993).
11. Hall, J. Uncovering hidden defects with neutrons. *Science & Technology Review*, May 2001, www.llnl.gov/str/May01/Hall.html.
12. d'Errico, F. Radiation dosimetry and spectrometry with superheated emulsions. *Nuclear Instruments and Methods in Physics Research B* **184**, 229–254 (2001).
13. d'Errico, F., Nath, R., Lamba, M. & Holland S. K. A position sensitive superheated emulsion chamber for three-dimensional photon dosimetry. *Phys. Med. Biol.* **43**, 1147–1158 (1998).

14. d'Errico F., Di Fulvio A., Maryański M., Selici S. & Torrigiani M. Optical readout of superheated emulsions. *Radiation Measurements* **43**, 432–436 (2008).
15. Bleuel, D. L. *et al.* Neutron activation analysis at the National Ignition Facility. *Reviews of Scientific Instruments* **83**, 10D313 (2012).
16. Thermo Scientific P 385 Neutron Generator,
www.thermoscientific.com/en/product/p-385-neutron-generator.html.

Acknowledgements. This project is supported by generous grants from Global Zero and the U.S. Department of State, funding from the Princeton Plasma Physics Laboratory, supported by DOE Contract DE-AC02-09CH11466, and in-kind contributions from Microsoft Research New England. The authors thank Francesco d'Errico, Jim Fuller, Duncan MacArthur, and John Mihalczo for valuable discussions and feedback. All simulations were run on Princeton University's High Performance Cluster.

The Princeton Plasma Physics Laboratory is operated
by Princeton University under contract
with the U.S. Department of Energy.

Information Services
Princeton Plasma Physics Laboratory
P.O. Box 451
Princeton, NJ 08543

Phone: 609-243-2245
Fax: 609-243-2751
e-mail: pppl_info@pppl.gov
Internet Address: <http://www.pppl.gov>