SAND2015-5117C

ZERO KNOWLEDGE PROTOCOL: CHALLENGES AND OPPORTUNITIES

Peter Marleau[1], Erik Brubaker[1], Nathan Hilton[1], Michael McDaniel[2], Richard Schroeppel[2], Kevin Seager[2], and Sharon DeLand[2]

[1] *Sandia National Laboratory, Livermore, California 94550, USA*

[2]*Sandia National Laboratory, Albuquerque, New Mexico, 87815, USA*

ABSTRACT

Glaser, Barak, and Goldston recently described an approach for nuclear warhead verification based on the cryptographic concept of a zero-knowledge protocol (ZKP). The verification approach included both procedural elements and a physical implementation. A group of Sandia National Laboratories researchers, whose expertise include radiation instrumentation design and development, cryptography, and arms control verification implementation, jointly reviewed the paper and identified specific challenges to implementing the approach as well as some opportunities. This paper compares the ZKP concept as presented by Glaser et al., to other warhead verification concepts and approaches described in the literature. The paper also summarizes challenges and opportunities for elements of this (new) approach. We note that ZKP as used in cryptography is a useful model for the arms control verification problem, but find that the direct analogy to arms control breaks down quickly. For example, fault tolerance in cryptographic ZKP is achieved by brute force repetition of the challenge/response procedure; in arms control, repeated measurements can be expensive. The ZKP methodology fits within the general class of template-based verification techniques, where template in this case refers to a reference measurement that is used for comparison. Template methods in general confirm that a given object is like another object that has already been accepted as a warhead by some other means. This can be a powerful verification approach, but requires independent means to trust the authenticity of the reference warhead – a standard that may be difficult to achieve. Note that this use of template differs from that of Glaser et al., who use the term to describe the accepted reference warhead. In the ZKP approach as described, the reference warhead must be present—with inspector confidence in its provenance—at each subsequent confirmation measurement. This imposes significant additional constraints on maintaining continuity of knowledge of the reference between inspections. Despite some technical challenges, the concept of last-minute selection of the pre-loads and equipment could be a valuable component of a verification regime. In addition, the neutron transmission image using a bubble detector array is an interesting idea that can be considered for template measurements independently of ZKP.

.

1.0. INTRODUCTION

A group of Sandia National Laboratories researchers, whose expertise include radiation measurement instrumentation design and development, cryptography, and arms control verification implementation, jointly reviewed the recent paper authored by Glaser, Barak, and Goldston that describes an approach for nuclear warhead verification based on the cryptographic concept of a zero-knowledge protocol (ZKP) (1) (2). Both challenges to the implementation and opportunities offered by the approach were identified.

It was noted that ZKP as used in cryptography can be a useful model for the arms control verification problem, but the direct analogy to arms control breaks down quickly. For example, fault tolerance in zero-knowledge cryptographic protocols is achieved by brute force repetition of the challenge/response procedure; in arms control, repeated measurements are expensive. The technical methodology presented by Glaser *et al.* fits within the general class of template-based verification techniques, where a reference measurement of the object's type is compared to a measurement. Template methods in general confirm that a given object is like another object that has already been accepted as a warhead by some other means. This can be a powerful verification approach, but requires independent means to trust the authenticity of the reference warhead – a standard that may be difficult to achieve.

Note that this use of template differs from that of Glaser *et al.*, who use the term template to describe the accepted, reference warhead itself rather than a measurement result. In the ZKP approach as described, the reference warhead must be present—with inspector confidence in its provenance—at each subsequent confirmation measurement. This imposes significant additional constraints on maintaining continuity of knowledge of the reference warhead between inspections.

In this paper, we provide the traditional cryptographic definition of the zero-knowledge protocol and compare it to the zero-knowledge implementation as presented in the paper by Glaser *et al.* We then discuss the reference warhead and the implications for authentication and chain of custody. We also discuss two interesting strengths of the protocol presented by Glaser et al., along with a few challenges and possible solutions.

2.0. SUMMARY OF THE PUBLISHED WORK

2.1. Definition of Zero-knowledge

In the cryptography context, Zero-knowledge has two properties as illustrated in the following scenario involving three parties, Alice, Bob, and John.

1: Alice and Bob want to confirm that Alice knows a certain piece of information, such as the solution to a math problem. Alice wants to do this without letting Bob learn anything about the solution, beyond simply confirming Alice's possession. Alice & Bob will exchange a series of messages to carry out the confirmation. If the communication is successful, Bob will be convinced with near certainty that Alice knows the solution, but he won't learn the content of the solution.

2: The nature of the exchange will be such that Bob cannot transfer his certainty to any third party (John). Bob can exhibit a transcript of his communications with Alice, but John cannot be sure that Bob didn't fabricate the whole thing. Bob believes Alice has the solution, since he knows he didn't make up the transcript, but that additional knowledge is intrinsic to Bob: the transcript alone is insufficient to prove that Alice has the solution.

As a concrete example, suppose there's a large number N, known to Alice, Bob, and John. Alice claims to know the prime factors of N are P and Q, but she doesn't reveal them. With the factors in hand, it's easy for Alice to compute square roots modulo N. Otherwise, it's very hard. Moreover, if Bob learns a few carefully chosen square roots, he can work out the factors.

Alice can confirm that she knows the factors by providing a sufficient number of square roots, but she can't allow Bob to submit his chosen numbers for root extraction, or he'll figure out P and Q. The dilemma is resolved by an exchange of a few messages, whereby Alice confirms that she can extract square roots, while Bob doesn't learn the square root of any number he wishes.

1. Bob chooses a number W, and computes $X = W^2 (mod\ N)$. He sends the challenge number X to Alice.
2. Alice chooses a random $R(mod\ N)$, and sends the numbers $R^2\ (mod\ N)$ and $X\ R^2(mod\ N)$ to Bob.
3. Bob chooses one of the numbers and asks for the square root.
4. Alice replies with R or $R\sqrt{X}\ (mod\ N)$.
5. Bob checks the answer.

If Bob cheats by just sending X without knowing W, he doesn't learn $\sqrt{X}$, since he doesn't know both R and $R\sqrt{X}$. If Alice is cheating, she could have faked either one of the square roots without knowing the factors of N, but she can't compute both without knowing $\sqrt{X}$. If Alice is faking, each challenge from Bob has a 50% chance of exposing Alice.

Zero-knowledge can be used to confirm logic gate results, and hence any computation, but it is computationally expensive. The Alice-Bob square root exchange can be completed in a few seconds on any modern computer, but a more complex result, such as Alice asserting that she knows a message M such that Hash(M) is a particular pre-specified value, can take hours or days of computation to carry out.

2.2. ZKP implementation as understood by the panel

Here we summarize an implementation of the ZKP discussed in (1). In the following list we outline the steps chronologically and highlight the responsibilities of the two parties involved, host and inspector. Attempts have been made to make this discourse agnostic to the specific confirmation technology wherever possible.

1. Template selection (optional) - The template is an agreed upon reference standard. The authors suggest that this could be a warhead selected from a deployed system in which case Chain of Custody (CoC) measures must be employed to maintain confidence. This is listed as optional as the authors state that differential measurements among a large number of objects in combination with some measure of provenance would lend some confidence in the results.
    a. Host – declares an object to be an authentic reference standard.
    b. Inspector's choice:
        i. Accepts declaration based on deployment status or provenance and CoC.
        ii. If agreed, confirms the authenticity of the object through another undescribed technical means or resigns to having no authentic reference standard.

2. Agreement of a maximum statistic in the measurement – The value of N_{max} is determined. This is the value of the mean number of counts in each observation value (the authors suggest that these could be the count values in each pixel of a fast neutron radiograph).
    a. Host – may be motivated to reduce the required measurement time.
    b. Inspector – may want to predict minimum number of counts required to remain sensitive to characteristics of "warheadness" depending on the technology used.
3. Generation of measurement preloads – Based on the agreed upon confirmation technology (the authors suggest fast neutron radiography), the inverse of a positive measurement on the template is determined in order to preload the instrument(s).
    a. Host – determines inverse of positive measurement and generates preload data sets such that, when combined with a measurement of the template, the result will be N_{max} counts in every measurement bin (image pixel) within statistical uncertainties (square root of N_{max} given Poisson statistics).
    b. Inspector – does not participate.
4. Presentation of declarations and preloaded instruments.
    a. Host – presents the inspectors with multiple declared treaty accountable objects, multiple preloaded instruments, and template(s) (if they exist).
    b. Inspector – selects any combination of objects and instruments which can/should include repeat measurements. Steps 5-7 below are repeated for each combination.
5. Confirmation measurement commences.
    a. Host – owns the instrument and conducts the measurement.
    b. Inspector – might be allowed to view the measurement to ensure that the selected combination is used, but not the preload data.
6. Measurement results confirmed.
    a. Host – probably owns and would want to check the result, but this is not discussed.
    b. Inspector – confirms that all counts are statistically consistent with N_{max} and thus the preloaded data was the inverse of the object measurement.
7. Post measurement.
    a. Host – may present the instrument to the inspector to confirm its authenticity.
    b. Inspector – may exercise the instrument to ensure that it was indeed functional during the measurement.

Two (or more) such confirmation measurements are performed on two (or more) objects. Because the inspector selects at the last minute which object is measured in each setup, the inspector gains confidence from a positive result that the two objects are the same, because if they were different there is only a 50% chance that each object was paired with the "correct" template. In the outlined protocol this is can be accomplished even without observing the measurements or confirming that the equipment and templates are indeed identical.

3.0. IS THIS ZERO-KNOWLEDGE?

Zero-knowledge uses statistical certification based on many tests which can only be answered by a party with the correct piece of information. Who has what knowledge at which step is critical to the process and this is where the ZKP analogy begins to fail. The inspector has nothing equivalent to a cryptographic key and therefore has no confidence in the result (other than proving that successful completion of all tests confirms that all items are statistically identical).

Some differences between an idealized implementation and measurements made in the real, physical world:

- Test outcomes in the real world are more ambiguous: In the crypto world, if Alice fails any of the challenges, Bob can say "I caught you, you're faking". The real world is more uncertain: a test could fail as false positive or negative with some probability. This isn't an unsolvable problem, but it means that more tests are needed to achieve a particular level of near certainty.
- It was noted earlier that tests are expensive (time-consuming), so the tradeoff between confidence level and test effort must be investigated.
- There's a tradeoff between information leakage and confidence level to consider.

The crypto protocol involves only two parties, plus a potential auxiliary player. We assume the two parties are opposed, with Alice's goal being to convince a skeptical Bob, who requires a solid test. In an arms control context, the parties may have more complex goals. Each party will have various incentives for the tests to succeed or fail, with various opportunities to alter the steps or influence the results.

As a simple example, the host might deliberately fail a test occasionally, either as a threat, or to maintain uncertainty. Recall that Saddam Hussein said he wanted his neighbors to think he had nuclear weapons, but believed the CIA would figure out that he didn't.

Another important difference is that third parties may actually want confirmation that Alice and Bob are not colluding to fake the results. The crypto-world requirement that Bob's knowledge is non-transferable is inverted: Both Alice and Bob want to confirm a successful test to third parties.

4.0. THE HIDDEN AUTHENTICATED STANDARD

As mentioned in the previous section, the protocol as outlined by the ZKP authors establishes a degree of confidence that the objects presented to the inspecting party by the host are statistically equivalent depending on the number of tests conducted. However, if the inspecting party also desires that they be statistically equivalent to a warhead, then a standard “pre-authenticated” warhead must also be included in the testing procedure.

The panel saw no way around this requirement given the approach outlined in the publication. Even if every item that the host has ever declared to be treaty accountable were found to be identical, they could also all be statistically equivalent to an empty container

The language presented in the publication was not clear in this regard, but we believe this is what was meant by the “template”. If this is true, then the authors suggest that an authenticated standard be presented as an object that could be chosen by the inspecting party at any time during confirmation. Presumably, the authenticated standard requires its own strict chain of custody measures in order to maintain continuity of knowledge of its authenticity.

The authentication problem arises again for initializing the standard warhead (template) into the regime. The authors do not address this other than to suggest that pulling the standard "template" warhead from deployment might lend confidence in its authenticity.

5.0. STRENGTHS

From the group's perspective there are two independent ideas presented in the ZKP paper: the role of inspector choice during the inspection/verification, and template matching with a physical information barrier. Because these two are presented together, they seem to be describing a single concept. However, they both have their own strengths and should be evaluated independently.

5.1. Inspector delayed choice

The aspect of this work that is most like cryptographic ZKP is the role of the inspector in the confirmation measurement process; the inspector is allowed to choose among detectors (preloads) and among objects to be measured after they have been prepared. As long as the measurement that is conducted is sensitive to relevant attributes of the object to be inspected, this can be done without consideration of the detection technology. It is this process that provides confidence to the inspecting party that the preloads and objects are statistically identical after several measurements are conducted. However, as discussed in Section 4.0, it only confirms that all of the objects are warheads to the degree that the "template" is believed to be one. This must be achieved through provenance; by previous confirmation and chain of custody or physical context, such as having been deployed on a delivery vehicle. It should be noted that the concept of obtaining confidence by making benchmarking measurements on randomly selected deployed items or with randomly chosen instruments has been suggested elsewhere (3) (4).

Though it was not emphasized in the article, the panel found that the role of the inspector in the measurement process was one of the more important aspects of the concept that the authors outlined. In the conventional concept of warhead verification, the host owns the measurement to protect their sensitive information. However, this results in a distinct lack of confidence in the outcome. By opening up the possibility for the inspector to have a role in the measurement process, namely inspector "delayed choice" (and possibly inspector provided test objects, see Section **Error! Reference source not found.**), there is much confidence to be gained even while keeping the detector, measurement, and data safely guarded.

5.2. Template Matching with several interesting advantages

Though the role of the inspector in the measurement process is most like its cryptographic namesake, it is the detection method presented that is more "zero knowledge" in implementation. As outlined in 2.2, the process of preloading a position sensitive detection plane with the expected inverse of a measurement allows a positive confirmation to be inferred from a result that theoretically contains no further information content.

Though the concept of preloading the detection system with an inverse of the expected measurement offers several interesting advantages, the panel felt that its implementation resembled something closer to template matching than to a Zero-Knowledge Protocol as is used in cryptography (Section 2.1). The "pad" that is preloaded into the neutron radiograph system is akin to a template; or more accurately, the inverse of a reference measurement. Successful certification results in a null or statistically flat result that theoretically reveals no information about the object.

In addition to this feature, the inverse template matching procedure offers several further advantages:

1. It becomes conceivable to carry out the entire confirmation measurement without the use of a digital device. Although, depending on the detection technology chosen, it still might be hard to avoid computers. For instance, a method to count or grossly integrate the number of bubbles in the proposed bubble detectors in an analog manner would need to be developed.
2. If a pre-authenticated standard object is among the objects that can be measured, then the inspector does not need to be involved nor keep CoC of the reference measurement used to create the pad. Though CoC of the standard would then be required.
3. The detection system is left in verifiable state. If a positive test results in a null, then the system could be transferred to the custody of the inspecting party to certify its functionality without risk of leaking sensitive information (however see Section 6.2).
4. Arbitrarily high confidence that the objects presented by the host are identical can be attained by making more measurements.

6.0. CONCERNS AND WEAKNESSES IDENTIFIED

Without making assumptions about any particular choice of detection technology or imaging scheme (active vs. passive), we have identified some concerning factors that may be weaknesses in the proposed verification scheme.

6.1. Fault Intolerance

In cryptography, fault tolerance in zero-knowledge protocols is achieved by brute force repetition of challenge/response; in arms control, repeated measurements may be expensive both in terms of time and capital. For each test, the host must preload a pad, conduct a measurement, read out the instrument (possibly in an analog fashion), and report the results. It is conceivable that the pads could be constructed in an automated fashion. However, depending on the radiation flux (whether passive or active) and the attenuation presented within the treaty accountable object, each measurement could take tens of minutes to hours to complete.

This could be offset by using multiple detection systems at the trade of capital costs. Even so, many more than tens of measurements per object are likely not conceivable. At such low numbers, one might worry about the perception of a negative result. It is likely that neither party will want to tolerate this possibility, so a different approach to fault tolerance is likely needed.

6.2. Potential sensitivity to information loss

Another kind of fault tolerance must be considered: even a single negative measurement outcome results in some potential loss of sensitive information from host to inspector. Any time a measurement does not result in a null, statistically flat, answer, some of the preloaded pad is revealed. This could be somewhat mitigated if the inspecting party was not allowed to keep a result, but merely view it. However, it is generally accepted that even the inspector's eyes should be considered the equivalent of handing over the information. The panel identified three potential avenues by which a negative could occur.

6.2.1. Object-detector registration

If perfect registration between the source, the object, and the detection system relative to the assumed alignment in the preloaded pad is not achieved, then any boundaries in an image would register as edges with excess counts on one side and deficit counts on the other. If a particular technology has sub-pixel resolution, as bubble detectors potentially do, then even very small misalignments can leave sensitive information within the pixels.

Even if the detector pixels are large with respect to misalignments, this information could accumulate in the relative rates of adjacent pixels over many measurements. This could be true even if the result of each measurement were statistically a null positive. After many measurements with a systematic bias, a statistically significant deviation could be obtained. If for instance, the detector counts were recorded or if the inspecting party was allowed to keep the detectors, sensitive information could gradually leak over the course of the inspection regime.

6.2.2. Precision of N_{max}

Similar to the misalignment problem, the dose delivered to each pixel (achieving N_{max}) must be precisely controlled. If, for instance, either the dwell time or the output rate of a neutron generator were over or under what was assumed in the preloaded reference, then the pad itself would be at risk. Take for example the extreme case that the dwell time was zero. In this case, the preloaded pad would be revealed in its entirety.

In the same way that small misalignments in registration could accumulate over many positive measurements, small imprecisions in arriving at the correct dose delivered (or dwell time) would achieve the same slow leak of sensitive information.

6.2.3. Verification of negative result

The panel felt that any verification system would be required to demonstrate the rejection of false items as a means to authenticate the equipment. However, as discussed in the previous two sections, given the procedure outlined in the paper, anything other than a true match would likely reveal sensitive information in the preloaded pad. Therefore, this can be seen as a weakness of the technique.

7.0. A FEW POSSIBLE SOLUTIONS

Potentially all of the points made in Section 6.2 could be resolved if spatial information is removed after the measurement; i.e. if all the detection pixels are dumped into a box and counted independent of their spatial location in the detector. Within the measurement scenario presented by the authors, all that is required to produce a negative result is for any single pixel to be statistically inconsistent with the agreed upon value of N_{max}. This can be determined independently of their relative spatial configuration.

Even if this approach is used, it is yet to be determined whether even the knowledge of the variance of counts, if statistically inconsistent with N_{max}, could reveal sensitive information. Therefore, in order to mitigate issues regarding the precise delivery of dose (Section 6.2.2), N_{max} might be monitored by an independent spectator detector. Alternatively, the host could verify N_{max} has been reached by sampling the measurement before revealing the result to the inspector. However if the host is allowed to withhold any of the measurement results, the advantages gained by the role of the inspector in choosing objects and pads are diminished.

If the detectors and the measurement results are owned by the host then the inspectors would not be able to aggregate and infer sensitive information over time if small deviations from a flat-field Poisson distribution were unavoidable.

As for the pre-authenticated standard warhead, the panel did not see an alternative. In order to have confidence that the inspected treaty accountable objects are not only statistically consistent with being identical, but also with being warheads, the authenticated standard must also be available to be chosen by the inspecting party.

8.0. SUMMARY AND CONCLUSIONS

In this work, the panel has outlined their understanding of the implementation of a zero-knowledge protocol as Glaser et al. proposed as a new approach to warhead verification in an arms control regime. Though it was generally agreed that this approach is not exactly a zero-knowledge protocol in the cryptographic sense, two related but separable innovative aspects were identified.

First, the inspector's last minute selection of which object is measured with which preloaded instrument provides confidence from a positive result that multiple objects are the same even without observing the measurements or confirming that the equipment and templates are indeed identical because if they were different there is only a 50% chance or less that each object is paired with the "correct" template.

The second innovation addresses the reality that, even without observing the measurement, the inspector must believe that a measurement is taking place and that it is sensitive to relevant characteristics of the object. The article proposes a fast neutron radiography measurement using an array of bubble detectors as the image plane. The detector array is preloaded such that a successful confirmation result is a uniformly populated array that reveals no sensitive information. Because the bubble detectors have a physical record of neutron interactions, this approach minimizes the importance of electronics in the measurement.

The panel reached a number of further conclusions, including:

- ZKP as used in cryptography is a useful model, but the direct analogy to arms control breaks down quickly.
- Last-minute inspector selection (analogous to a ZKP challenge) could be a valuable component of a verification regime, reducing requirements on e.g. equipment authentication.
- ZKP is in the category of template measurements, i.e. it confirms that a given object is like another object that has already been accepted as a warhead by some other means.

- In the ZKP as described, that accepted object ("template") must be present—with inspector confidence in its provenance—at each subsequent confirmation measurement.
- In cryptography, fault tolerance in zero-knowledge protocols is achieved by brute force repetition of challenge/response; in arms control, repeated measurements are expensive, so a different approach to fault tolerance is needed.
- The neutron transmission image using a bubble detector array is an interesting idea that can be considered for template measurements independently of ZKP.

Finally, the panel has identified and suggested several topics for ongoing research that would address some of these concerns and gaps.

9.0. REFERENCES

1. **Glaser, A., Barak, B., Goldston, R.** A zero-knowledge protocol for nuclear warhead verification. *Nature.* 13457, June 26, 2014, Vol. 510, pp. 497-502.
2. *A New Approach to Warhead Verification Using a Zero-Knowledge Protocol.* **Glaser, A., Barak, B., Goldston, R.,.** Orlando, FL : s.n., 2012. INMM 53rd Annual Meeting.
3. **Engineering, Arms Control and Nonproliferation Technologies Office of Nonproliferation and.** *Technology R&D for Arms Control.* s.l. : U.S. Department of Energy National Nuclear Security Administration Defense Nuclear Nonproliferation, 2001. pp. 4-5.
4. *Random Selection as a Confidence-Building Tool.* **MacArthur, D., Hauck, D., Langner, D., Smith, M., Thron, J.** 2010. Annual Meeting of the Institute for Nuclear Materials Management.