LA-UR- 10-03785

*Approved for public release;*
*distribution is unlimited.*

| | |
|---|---|
| *Title:* | Defining the Questions: A Research Agenda for Nontraditional Authentication in Arms Control |
| *Author(s):* | Danielle K. Hauck, Duncan W. MacArthur, Morag K. Smith, Jonathan Thron, and Kory Budlong-Sylvester |
| *Intended for:* | 51st Annual Meeting of INMM |

Los Alamos
NATIONAL LABORATORY
EST. 1943



Form 836 (7/06)

# Defining the Questions: A Research Agenda for Nontraditional Authentication in Arms Control

Danielle K. Hauck, Duncan W. MacArthur, Morag K. Smith, Jonathan Thron, and Kory Budlong-Sylvester

Los Alamos National Laboratory, Los Alamos, NM 87545

## Abstract

Many traditional authentication techniques have been based on hardware solutions. Thus authentication of measurement system hardware has been considered in terms of physical inspection and destructive analysis. Software authentication has implied hash function analysis or authentication tools such as Rose. Continuity of knowledge is maintained through TIDs and cameras. Although there is ongoing progress improving all of these authentication methods, there has been little discussion of the human factors involved in authentication. Issues of non-traditional authentication include sleight-of-hand substitutions, monitor perception vs. reality, and visual diversions. Since monitor confidence in a measurement system depends on the product of their confidences in each authentication element, it is important to investigate all authentication techniques, including the human factors.

This paper will present an initial effort to identify the most important problems that traditional authentication approaches in safeguards have not addressed and are especially relevant to arms control verification. This will include a survey of the literature and direct engagement with non-traditional experts in areas like psychology and human factors. Based on the identification of problem areas, potential research areas will be identified and a possible research agenda will be developed.

## Introduction

The tracking and verification of special nuclear materials (SNMs) plays an important role in global security initiatives. Future arms control treaties may require inventories of nuclear warheads and tracking of SNM during the dismantlement process. The International Atomic Energy Agency (IAEA) is charged with monitoring enrichment and reprocessing facilities to ensure that SNM is not diverted for non-peaceful use.

Well-established technologies exist for detecting and characterizing SNMs. However, the measurement results may contain sensitive information (from either a security or a commercial perspective) that the host country would like to protect. Protected information may include the shape of weapons components, the isotopic ratios of weapons materials and proprietary technologies. Measurement systems must be placed inside a physical and/or technological enclosure, often called an information barrier, to protect sensitive information.

In this context, authentication is a confirmation that the measurement system used for treaty verification is genuine and works according to agreed standards. The presence of an information

barrier limits the amount of information available to the monitor and complicates the task of authentication. In many measurement systems the information barrier is a combination of physical enclosures, electromagnetic shielding and signal processing design. Only agreed upon measurement results are allowed to pass through the information barrier.

There are expected to be administrative controls utilized to protect information in addition to the technological components of the information barrier. This leads to several assumptions regarding the use of these measurement systems and implementation of the verification process. These include:

1 The host country will supply and handle the verification measurement system,

2 Any systems or system components that have recorded classified information will not be allowed to leave the host country or be analyzed by the monitoring party, and

3 Inspectors may observe and give instructions to members of the host country as part of an agreed-upon procedure.

Certain aspects of authentication have been well developed. For instance, hardware and software components can be authenticated through reverse engineering, hash functions and destructive analysis. Also familiar to authentication is the use of Tamper Indicating Devices (TIDs) to reveal possible adverse handling of instruments or components.

Several aspects of authentication are visited in other disciplines and even in everyday life. Security questions for online banking, internet data certificates and the secret password given to children are all forms of authentication. The authentication community would benefit from exploring the insights offered by experts outside of the field of nonproliferation. In particular we should consider the capabilities and weaknesses of the latest data encryption and data authentication technologies. The effects of human factors in the authentication process have not yet been examined. How might the host country utilize methods such as distraction or sleight-of-hand to discredit the authentication process? Can typical human responses to confusing or ambiguous situations be predicted? Are there formal means of assessing the degree of uncertainty contributed by human factors?

To develop these questions we held an Internal LANL workshop on May 27, 2010 titled Authentication Challenges in Treaty Verification. We invited two speakers to introduce the problem of authentication from the perspectives of treaty verification and enrichment plant safeguards. We invited three speakers to respond to the problem of authentication from their own discipline or perspective. The speakers included Todd Conklin on Organizational Psychology, Beth Nordholt on Quantum Cryptography and Stephan Eidenbenz on Zero Knowledge Protocols. A total of 25 people attended the workshop from wide ranging disciplines at LANL to discuss the problem and a research plan for the future.

In this paper we use the discussions during the workshop, individual discussions with subject matter experts (SMEs) and scientific literature to identify the primary authentication problems in the context of nuclear nonproliferation that have not yet been satisfactorily addressed.
We will discuss future work that focuses on individual problems that were identified and suitable technological solutions. The focus of this paper is on the least well-developed problem; human

factors in authentication. However, we also touch on data encryption techniques and other technical solution options that have not been discussed in the context of authentication.

**Workshop Themes**

There were several recurring topics that arose while discussing authentication in the workshop. These ideas speak to the general needs to have in mind while moving forward with authentication.

*Who must be convinced?*
Many subject matter experts at LANL and other laboratories have considered the problem of authentication. While these scientists must be confident in the quality of authentication, this confidence must be transferred to the political leaders entrusted with national security.
So far, authentication has been discussed most frequently in the context of a bi-lateral or tri-lateral initiative. In this context each country, and possibly the neutral third party must be confident in the authentication. However, a more far reaching, but important, goal is to give confidence to other countries of the world. In addition, future treaty initiatives will hopefully include multiple countries instead of two or three.

The focus of authentication is often placed on preventing false negatives, or the failure to correctly identify a weapon. However, authentication should also prevent against the false detection of a weapon. False positives could result in lash backs by local interest groups. The administrative procedures for handling this scenario are outside the realm of this paper. However, this concern brings to mind that the confidence gained by authentication can be important to local citizens as much as it is to country's leaders.

*What kind of information should we draw on?*
Authentication has been considered a technological problem so far and accordingly most proposed authentication methodologies are technological in nature. However, much more information is available to draw on. For instance, the credentials and past history of the scientists handling the measurement system, the general appearance and behavior of the people present and cultural tendencies should all be taken into consideration. Cultural knowledge can help predict the risk of nations, groups or individuals attempting to cheat. Knowledge of global scale and local scale politics can help address who has the motivation to cheat. This kind of knowledge has been integrated to predict dangerous possible sources of proliferation. However, it has not yet been utilized as a whole in the context of authentication. Complete confidence will not come from any one authentication method. Instead, good confidence will come from the combined results of several methods, both technological and sociological.

*What level of confidence can be expected?*
It is the task of experts to provide authentication options along with a description of their cost, difficulty to perform and anticipated efficacy. This requires some evaluation of the degree of confidence granted by each method of combination of methods. Any evaluation should include the possible effects of human error or human deception as well as the strict statistical uncertainties in the technology utilized.

**Traditional Authentication**

Many tools are proposed for authentication in arms control scenarios including:

- *Reverse engineering* – complete (possibly destructive) analysis and validation of hardware. This refers to the use of hash functions in the case of software. This is very robust and reliable. The problem comes from the inability to apply it to measurement instruments after they have been used by the host country.
- *functional testing* – complete and convincing testing of a measurement system to verify correct and expected operation, possibly with the use of authenticated standards. If feasible, this is the strongest form of authentication.
- *cooperative design* – Cooperative and open-knowledge design and construction of the measurement system between 2 or more participating countries. The detailed knowledge gained would make functional testing easier.
- *design transparency* – Full documentation of the host-built AMS given to the inspecting party. This will be more time and cost intensive than the option of cooperative design, since the documentation must be understood after the fact.
- *continuity of knowledge* – ability to maintain a record of an object and lack of tampering for a period of time and possibly between two points in space. This has applications throughout authentication, including on measurement instruments after they have been functionally tested and on identical instruments before they have been analyzed.
- *random selection* – ability for the inspecting party to randomly select components to be used in the measurement system and to be destructively analyzed by the inspecting party. This technique is important if the actual measurement instruments are off limits and analysis must be performed on identical components.

Authentication will require the integration of all of these techniques. An important common element to functional testing, reverse engineering and validation techniques is the need for the monitor to have confidence that the approach is being applied to either the actual component used in the measurement or an equivalent component. If the actual component is off limits, the monitor must be able to trust that an approach like random selection provides access to a true equivalent to the actual component.

The concept of random selection can be thought of as four steps: (1) The host presents several "identical" copies of a component or system to the monitor. (2) One (or more) of these copies is randomly chosen by the monitors for use in the measurement system. (3) Similarly, one or more is randomly chosen to be validated further at a later date in a monitor-controlled facility. (4) Because the two components or systems are identical, validation of the "validation copy" is equivalent to validation of the measurement system. This procedure sounds straightforward, but effective application may be quite difficult. Although random selection is often viewed as a panacea for confidence building, the amount of confidence generated depends on the monitor's continuity of knowledge for both validation and measurement systems. [1]

Continuity of Knowledge (CoK) is a need that runs throughout nonproliferation. For example, a very strong version of authentication would be complete functional testing of the measurement system. However, CoK is needed to prevent tampering between the time of authentication and the actual use in measurements. Functional testing will likely require the use of authenticated standards. CoK must be maintained on the standards to ensure true testing of the device. In the case of random selection, CoK is needed on the components from the time they are chosen to the time they are used in the measurement or analyzed with reverse engineering. Tamper Indicating Devices (TID) are commonly used for Continuity of Knowledge (CoK) concerns.

Cooperative Design may supply a good opportunity to build trust and establish the mutual benefits of authentication. An important aspect of deterrence is to reduce or eliminate the motivation to cheat. The scientists and politicians in other countries do not necessarily believe that authentication is in their best interest. It is possible that cooperative design would help establish a mutual feeling of the beneficial nature of authentication. In addition, with the potential for future multi-party treaties, each country should feel treated equally. This might best be accomplished by cooperative design and the mutual establishment of a measurement system and protocol to be used in all countries. There are several difficulties associated with cooperative design that should be acknowledged ahead of time. The practicalities of such a venture would require patience, understanding, time and money. Some of the benefits of cooperative design might be gained from transparent design.

A major component to the success of the authentication tools is human factors. The authentication process cannot be treated as a purely technological system. The importance of human factors has been neglected so far in the authentication community. While human factors such as the ability to be distracted or deceived can be used to undermine authentication, the consideration of cultural presets and human behavior can be used to strengthen the authentication effort. Human factors must be considered in any evaluation of the effectiveness of authentication.

**Human Factors**

From a psychological perspective the problem of authentication represents a goal conflict. The inspecting party would like to acquire as much information as possible and the host party would like to reveal as little information as possible. In fact, each party suffers the conflict individually. A country would like to obtain as much information as possible from other countries while revealing little of its own. This conflict leads to a natural potential for deception.

Todd Conklin is an Organizational Psychologist at LANL with over 20 years of experience in human behavior. Figure 1 demonstrates the parallel structure envisioned by Todd. The Sleight of Hand side represents the magician problem, the ability to be distracted and have a component switched without the monitor's knowledge. Operational restrictions encompass any kind of organizational component that can be used to influences an inspector's choices.

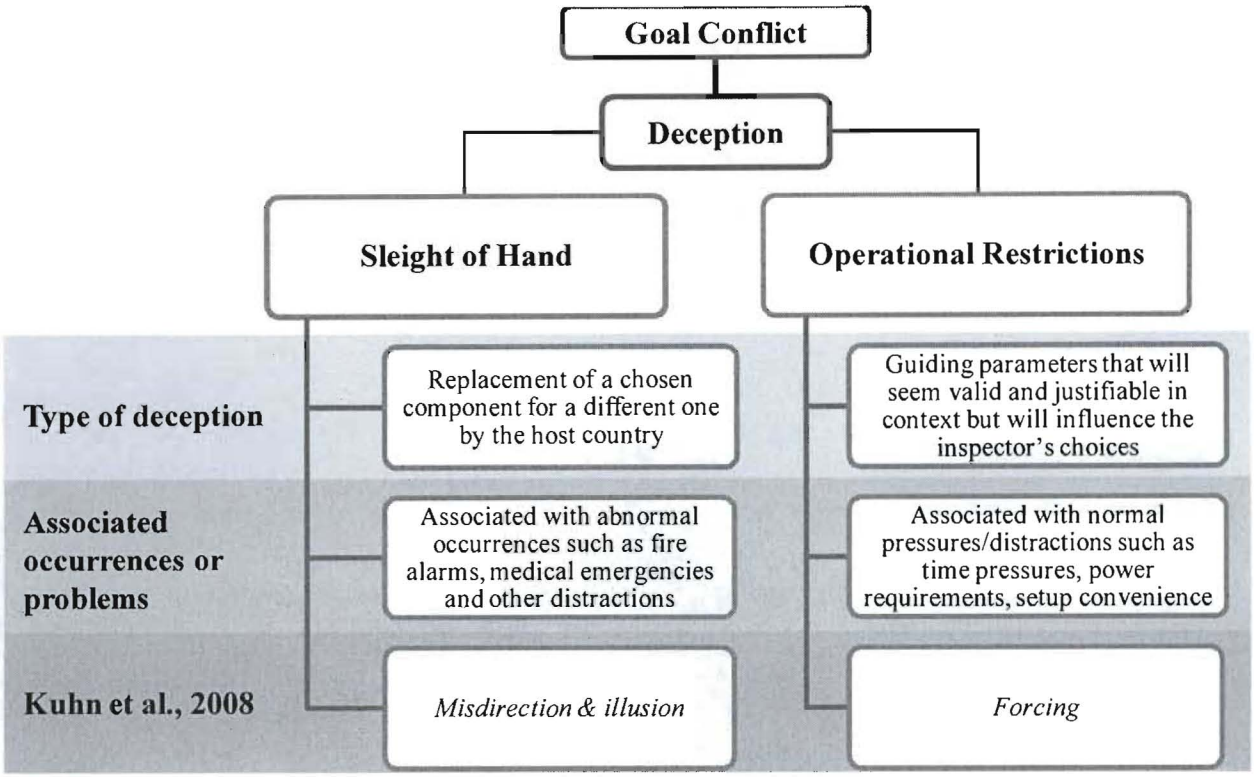


**Figure 1. Human Factors in the context of authentication and treaty verification.** The situation can be regarded most basically as a goal conflict in which the inspecting party would like to acquire as much information as possible and the host party would like to reveal as little information as possible. Kuhn et al. (2008) discussed similar effects with different notation.

*Sleight of Hand*

As information recording devices, humans are flawed. Humans have been optimized to filter the information that they are consciously aware of. This makes it easier to respond to potential threats. However, this results in conscious awareness of certain items or events to the exclusion of other information that is entering our brain.

The phenomenon of change blindness is that humans do not notice changes in objects unless they are consciously attentive to the object.[2] The human brain is only able to register one movement, or change, at a time. In addition, we do not notice change unless we register the associated movement. As a result, it is possible to switch out objects or people without a person noticing if the associated movement is hidden or if the person's eye is registering a different movement at the time. The effects of change blindness are most extreme when the change event is not expected. Most people do not have an intuitive understanding of their own limitations in this regard and believe that they would notice such changes.

The phenomenon of change blindness brings into question the amount and quality of information that is being consciously registered. In some cases, the human brain compensates for lack of information by filling in knowledge gaps. Our brains predict the result of eye movements and update the visual image before information from the new image is processed.[3] Crime witnesses with poor observation of an event will retroactively supplement their memory of the event with information obtained at a later time. [4].

With these considerations comes the concern that inspectors can be purposely deceived as to the events that they are actually witnessing. This problem is relevant to the random selection methodology, but can be a potential threat in almost any inspection scenario. In the random selection scenario, the "measurement" component may be switched before installment in the detector system, or the "take-home" component may be switched prior to leaving the facility.

Kuhn et al identified three areas, misdirection, illusion and forcing, used by stage magicians to perform tricks.[5] Misdirection and illusion can be used to deceive an inspector about what they are observing. Misdirection makes use of the fact that humans do not have visual perception of objects or events that they are not attentive to. Magicians can control the locale of a person's attention with stimuli such as movement, high contrast and novelty. People tend to look in the same direction as the magician. Magicians can use body language to set up a certain level of attentiveness or laxness in the observers. They also utilize the fact that people are less likely to see something that they don't expect. Therefore, multiple repetitive innocent actions will reduce people's expectation of a deceptive change. Illusion takes advantage of the fact that the brain extrapolates motion to account for the ~100ms delay between visual stimulus and conscious perception. If the extrapolation is wrong, it is possible for the brain to perceive seeing something that never happened.

Some ways to combat sleight-of-hand is to place any components in to TIDs as soon as they are selected. There should be well-defined boundaries of interaction between the inspectors and the hosts. There should not be intricate interactions or conversations in which the inspectors may subliminally influence the hosts through speech patterns or body language. The inspectors should be hyperaware of the focus of their attention and times that they feel distracted. Any kind of "abnormal" occurrence can serve as a distraction for switching components. Since the ability to perceive is so strongly linked with conscience attention, sleight of hand would be associated with abnormal occurrences such as fire alarms, birthday cakes, medical emergencies and other distractions.

*Operational Restrictions*

As an organizational psychologist, Todd Conklin brought to our attention the potential for using operational restrictions to influence the inspectors through a series of functional preferences which seem (and may be) valid needs. The occurrence of this type of deception is associated with normal pressures and distractions, such as time pressure, power requirements and setup convenience. People are predictable in their decision making and the organization of the random selection process can greatly influence inspectors. Under time constraints people tend to simplify their selection process.[6] Fewer criteria are considered when making a decision, so time constraints can amplify the effects of organizational restraints.

This type of magician trick was also discussed by Kuhn et al. (2008) as forcing. In this case a magician manipulates the presentation of a group of objects to favor a particular decision. For example, a person is more likely to choose an object that they have had longer exposure to. Magicians also use time pressure to increase the person's susceptibility to the preferred choice. Afterwards people retroactively make up reasons to rationalize the decision that they made.[7] This phenomenon, called choice blindness, means that inspectors may not realize that they were influenced, even if they are presented with an item or outcome that is different from the one that they selected.

Time pressure is a common problem which often occurs in the work place. The important and sensitive nature of the inspector's task may contribute to a perception of time pressure. Time pressure can be used to make some components more preferable to an inspector through ease of accessibility. Inspectors should be taught that this kind of motivation is very predictable in humans. They should be taught ways to ensure that they choose a representative sample from the group. For instance, do not choose the two components that happen to have the longest power cords just because they may be the most convenient. One person should never be responsible to make the decision on their own. Sample components should be selected in groups. For instance, if one component is needed for the measurements and one is needed for destructive testing, both components should be selected at the same time. The component for testing should be specified after both components have been chosen. On occasion, these components should be switched at the last minute. Other options of using random selectors (dice, a coin or a random number generator) have been suggested for ensuring that the inspectors select a representative sample.

Cultural differences will play a role in how selection guidelines are played out. Americans are very individualistic in their social interactions and Japanese have finely prescribed social protocols based on context. Russians can be informal at times and have strong social expectations at other times. For instance, heavy drinking and socializing is common in Russian. The inspectors will have to balance this expectation while keeping an appropriate distance to minimize the possibility of influence.

Personal differences between inspectors must be carefully managed to ensure that decision making is always made as a group. Inspectors should be peers and not have authority over each other for the same reason. Inspectors may be required to take a personality test in order to understand each other's reactions in an attempt to minimize the effects of conflict while on an inspection mission. Inspectors should have a safe place in which to communicate away from the supervision of the host country.

**Future Work**

It is evident that more work is needed on the effects of human factors in authentication. Many of the suggested authentication procedures assume that the monitor is able to be a perfect observer and to make free choices. Both assumptions can be undermined by basic human perception skills. It is necessary to focus future research on the effects of human factors and ways to protect against them. We must identify the ways that a host can influence choice to undermine the free

will of the monitor, and protect against it by identify ways to identify influences and limit their effectiveness.

Cooperative Design of monitoring equipment has benefits in creating trust, educating developers and creating advocates for the mutual benefits of safeguards and authentication. The degree of these benefits should be studied more fully. Given the practical difficulties, the net benefit must be determined. The practical difficulties in a possible multi-lateral situation should be considered. Is there a way to move to the transparent design process and still maintain most of the benefits offered by cooperative design.

The double-blind experiment was suggested as a possible procedural check against deception. In an anticipated treaty verification, each item may be checked in a go/no-go manner. The possibility of deception is decreased if neither the monitoring party nor the host party are immediately aware of the outcome of each measurement. Encoding the results of the measurement makes it harder for the host party to anticipate the type of signal that should be simulated in a deception scenario. There may also be a behavioral tendency to reduce the motivation to deceive when the results are not immediately known. This kind of procedure would present some practical difficulties in an inspection scenario, but the possible benefits should be considered more fully.

An important task of the authentication community is to supply effective options to decision makers and political leaders. It will not be possible to make use of all authentication options due to time and political constraints. Therefore, it is imperative to evaluate and quantify as much as possible the effectiveness of methods and combinations of methods. It is necessary to combine technological assessments of measurement performance with sociological and psychological assessments of protocols and organizational systems. This points to a need for a consistent method of combining many different types of errors. We hope to work with scientists at LANL to develop a model for complex systems that can combine many types of errors. Such a model will also be capable of predicting best paths forward which will result in the greatest reduction of uncertainty over all.

Quantum key cryptography may supply a robust method for security key generation which may be useful in maintaining CoK on equipment. However, the ability to "zero" or "scramble" information may be even more relevant. If actual measurement equipment is off-limits, then authentication relies on the equivalence of the validated equipment to the actual measurement equipment. Random selection has been cited as a possible method for ensuring equivalence. However, the ability to "zero" information, utilized by cryptographers might be useful for scrambling the information in measurement equipment so that it can be released to the monitor.

Zero knowledge protocols presents an interesting approach to proof of knowledge without revealing the actual information. In the authentication context the goal is to confirm that the item contains nuclear material with certain attributes without revealing the measurement information. It may be possible to apply zero knowledge protocols by writing queries that are sensitive to the accuracy or outcome of a measurement. In zero knowledge protocols, a series of queries are applied until the desired confidence is achieved.

## Conclusions

An Internal LANL workshop on Challenges in Treaty Verification was held in Los Alamos, NM on May 27, 2010. The purpose of the workshop was to make use of expertise at LANL that was previously untapped for measurement authentication in the context of treaty verification. The workshop demonstrated that this outreach effort was a needed first step, but that further incorporation of technological and sociological solutions was needed. The authentication process cannot be treated as a purely technological system. While human factors such as the ability to be distracted or deceived can be used to undermine authentication, the consideration of cultural presets and human behavior can be used to strengthen the authentication effort. Human factors must be considered in any evaluation of the effectiveness of authentication.

Multiple technologies and procedural approaches were also discussed and many of them warrant future work. The benefits of cooperative design, especially as it relates to human factors, should be evaluated more fully. The double-blind experiment may provide an effective procedural means to reduce the ability and tendency to deceive. Quantum key cryptography and zero knowledge protocols may provide useful tools as well as methodologies for authentication.

The effects of human factors as well as the discussed technological and procedural solutions will be pursued more fully in future work.

## Acknowledgements

The Internal LANL Workshop on Challenges in Treaty Verification was sponsored by the Nonproliferation and International Security Program Office. The authors would like to thank all of the participants of the Workshop for helping us to approach authentication with a holistic perspective, especially including the speakers Brian Boyer, Todd Conklin, Stephan Eidenbenz and Beth Nordholt.

## References

1. MacArthur et al (2010) Random Selection as a Confidence-Building Tool.
2. Simons, D. J. and R. A. Rensink (2005) Change blindness: past, present and future
3. Duhamal, J-R., C. L. Colby and M. E. Goldberg (1992) The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements. Science 255, 90-92.
4. Busey, T. A. and G. R. Loftus (2006) Cognitive Science and the Law
5. Kuhn, Gustav, Alym A. Amlani and Ronald A. Rensink (2008) Towards a science of magic.
6. Benson, L. and L. R. Beach (1996) The effects of time constraints on the prochoice screening of decision options
7. Johansson, P., L. Hall, S. Sikstrom and A. Olsson (2005) Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task. Science 310, 116-119.