Analyzing Vulnerability Results for Tags and Tamper-Indicating Seals

Roger G. Johnston and Anthony R.E. Garcia

Vulnerability Assessment Team
Los Alamos National Laboratory
MS J565, Los Alamos, NM 87545

Keywords: tamper detection, vulnerability assessments,
security, tampering, design of experiments

Introduction

Tamper-indicating devices, often called "seals", are meant to detect unauthorized access, entry, or tampering (NFESC, 1997; NFESC, 2000; Johnston, 1997c). Seals are widely used for a variety of applications including access control, cargo security, pilferage detection, banking, courier services, document and records integrity, customs, law and drug enforcement, hazardous materials accountability, nuclear safeguards & nonproliferation, counterespionage, counterterrorism, and consumer protection (Johnston, 2001d; Tyska, 1999). The U.S. Army frequently uses seals to detect pilferage and tampering with weapons during storage and shipment, and also to secure ammunition, medical supplies, soldier's personal property, courier packages, and classified documents.

Unlike locks, seals are not designed to resist, complicate, or delay unauthorized access. Instead, they record that it took place. Also unlike locks, seals must be inspected, either manually or electronically, to do their job. Seals differ from intrusion detectors ("burglar alarms") in that unauthorized access or entry is not reported immediately. This has both advantages and disadvantages (Johnston, 2001c).

Seals take a variety of forms. They can be frangible foils or films, plastic wraps, pressure-sensitive adhesive tapes, crimped cables or other (theoretically) irreversible mechanical assemblies; security containers or enclosures that give evidence of being opened; devices or materials that display irreversible damage or changes when manipulated; and electronic or electrooptic devices and systems that continuously monitor for changes, such as a break in an electrical cable or fiber-optic bundle. Perhaps the most familiar everyday example of seals is the tamper-evident packaging found on over-the-counter pharmaceuticals.

A tag is a device, or an applied or intrinsic feature, used to identify an object or container. A familiar example of a tag is the license plate on a car. When used for security purposes, a tag should be difficult or expensive to counterfeit, as well as difficult to lift. To "lift" a tag means to remove it from one object and attach it to another without damaging the tag and without being detected.

Tags and seals are related in that an effective security tag must be able to detect tampering. An effective seal, in turn, must have a unique (tag-like) characteristic or "fingerprint", such as a serial number. This is necessary so that it is not trivial to remove the seal from an object or container and replace it with a duplicate.

The Vulnerability Assessment Team at Los Alamos National Laboratory (LANL) is involved in trying to improve tamper detection for a variety of different applications, commercial as well as government. We conduct vulnerability assessments on tags, seals, and tamper detection programs, develop new tags and seals, and work on ways to improve existing tamper detection methods (VAT, 2002).

A "vulnerability assessment" (Jones, 1996; Johnston, 1997b) involves discovering and demonstrating ways to defeat security devices, systems, or programs. It may also involve suggesting countermeasures. To "defeat" a seal means to open it, then reseal (using either the original seal or a counterfeit) without being detected. Similarly, to "defeat" a tag means to counterfeit or lift it, without being detected. "Attacking" a tag or seal involves undertaking a sequence of actions designed to defeat it.. A successful attack is also called a "defeat".

There are two aspects of vulnerability assessments on tags, seals, and tamper detection programs that are particularly tricky: (1) designing effective vulnerability assessment experiments and (2) analyzing and reporting the results. This paper examines some of the complications, problems, and issues associated with these matters.

Complications, Problems, and Issues

Quite a number of different complex factors affect the design of experiments for testing tags, seals, and tamper detection. These factors and others also greatly complicate the statistical analysis and reporting of results.

One of the major problems in performing vulnerability assessments is that the field of tamper detection is not well developed. Although tags and seals have been used for at least 7,000 years (Johnston, et al., 2001b), they are poorly understood (Johnston, 2001d). There is little in the way of formal theory, few meaningful standards for performance or testing, and considerable confusion among end users about how to use tags and seals effectively (Johnston, 2001a). We have observed considerable misunderstanding among security professionals about tamper detection concepts, strategies, and terminology. There is frequently, for example, confusion about the difference between locks, seals, and tags. Expectations for tamper detection are often vague or unrealistic.

In our experience, there is also often a lot of confusion and wishful thinking associated with vulnerability assessments. Many security managers believe that a vulnerability assessment should ideally find no vulnerabilities. Our view is that multiple vulnerabilities are always present in ANY security device, system, or program. Discovering some of these vulnerabilities provides the opportunity to mitigate or eliminate them. Thus, the discovery of vulnerabilities should be viewed as good news, not bad news. Indeed, a vulnerability assessment that finds no problems has zero value.

Security managers (or their supervisors or auditors) often feel that if countermeasures and recommendations arising from a vulnerability assessment are implemented, this is an admission that they have been negligent or incompetent in the past. Vulnerability assessors must make allowances for such a mindset, and attempt to counter it. This can be quite a challenge.

A related problem that often plagues vulnerability assessments is the prevalence of absolutist ideas about security. To many people, a security device, system, or program is either secure, or it has vulnerabilities and is insecure. In reality, security is a continuum. Nothing is either fully secure nor completely insecure. A binary view of security is both unrealistic and dangerous (Johnston, 2001a).

Unfortunately, we in the Vulnerability Assessment Team are all too familiar with another problem with vulnerability assessments: "Shoot the Messenger Syndrome". It is all too common when security problems are discovered for the vulnerability assessors (often called "black hatters"!) to be viewed as the problem, rather than the vulnerabilities themselves. The Nobel prize-winning physicist Richard Feynman has written amusingly about this phenomenon (Hutchings, et al., 1985). Sometimes security programs are evaluated be personnel who may damage their own careers if they report significant or numerous security problems. Effective vulnerability assessments rarely occur in such an environment.

In many cases, security managers and supervisors, or manufacturers and vendors of security products do not want vulnerability assessments done because they highlight problems. This attitude is not conducive to effective security. Because security professionals (and security programs) are often judged by the lack of problems, however, this situation is difficult to avoid.

Often, vulnerability assessments are conducted by personnel who have a serious conflict of interest. It is not unusual to find that a security product or system has been analyzed and tested by proponents, manufacturers, or vendors of that product or system. Not surprisingly, few vulnerabilities are typically found in such analyses. Sometimes, vulnerability assessors are chosen who are clearly unqualified or unimaginative. They also tend to uncover few, if any, vulnerabilities.

A problem that we have frequently encountered in conducting vulnerability assessments is that the use protocol for a given tag or seal may be poorly defined, inconsistent, or not formalized. The "use protocol" is exactly how a tag or seal is used. This includes the entire lifecycle of the tag or seal including procurement, shipping, storage, checkout, installation, inspection, removal, disposal, data handling, analysis, interpretation, postmortem forensics (if any), and training of security personnel. A tag or seal is no better than its use protocol (Johnston, 1977a). Discovering and demonstrating defeats is difficult if it is not clear what use protocol must be defeated.

The human factors associated with defeating tags and seals are particularly tricky to handle. While defeating a lock, safe, or vault (for example) is mostly about beating hardware, defeating a tag or seal is primarily about fooling people. That psychological factor can be difficult to model, predict, or analyze. It can also be quite difficult to model and predict human error, yet human error is responsible for most security failures.

One frustrating problem with vulnerability assessments is that tags and seals are often exposed to a vulnerability assessment only after the design is finalized and the product is in production. By then, it is usually too late to make any changes in the design that might mitigate or eliminate vulnerabilities. Ideally, vulnerability assessments should be conducted on a tag or seal iteratively, throughout the various design phases of the product. This has the additional advantage of make the vulnerability assessors part of the design team, rather than the "enemy"—thus increasing the chances that their warnings and recommendations will be heeded.

Another very common problem with conducing vulnerability assessments on tags and seals is the paucity of samples made available to the assessment team. We have often been asked to find the vulnerabilities in a seal when we have been given exactly one sample—which must be returned to its owner undamaged. The most effective vulnerability assessments require dozens, if not hundreds of seal samples. Some, though not all, of the test seals usually need to be destroyed during the assessment process. It is usually not reasonable, in our view, to assume that an adversary won't have access to large numbers of seals for testing and practice—especially since most seal manufacturers give away samples for free.

Designing and executing a vulnerability assessment on a tag or seal is often constrained by time and funding. Adversaries who might try to attack tags or seals may not be so constrained. A time- or budget-limited vulnerability assessment, moreover, requires some kind of prioritization of the hundreds of possible attacks. Time and money will usually not be available to study them all. Not all possible attacks will be relevant or ultimately prove to be successful, and some of the attacks that do ultimate work may end up consuming more time and money to develop than they are worth. There are usually no obvious guidelines for how to prioritize attacks, though experience seems to be helpful.

A related complication is that we don't automatically know when the best attacks have been found. The best attacks may forever go undiscovered, or be discovered only at a later date by a different set of vulnerability assessors (or adversaries). Vulnerability assessments thus have no clear-cut end point, and so it is never clear when the "experiment" is over.

It can be difficult in analyzing vulnerabilities to know what adversaries might attempt. The exact identity of all possible adversaries, and the resources/capabilities available to them are usually unknown and can only be speculated about.

There are often "recursion" problems with vulnerability assessments. By this we mean problems associated with iteration and with hitting a moving target. Often, for example, after being shown a serious seal vulnerability, a distressed security manager will ask if there isn't a simple countermeasure. Usually there is. Once reassured that the vulnerability can be easily dealt with, the security manager will relax—yet never actually implement the countermeasure! Another example of recursion occurs when the recommended countermeasure, whether a change in the seal use protocol or to the seal design itself, introduces new problems and vulnerabilities. It is very difficult to fully foresee in the original vulnerability assessment, all the vulnerabilities associated with a theoretical tamper detection regime that may come into existence as a result of implementing some or all of the initial recommendations. Ideally, a new vulnerability assessment should be conducted after recommended changes are actually made, but this is rarely done.

"Compliance mode" is another problem that threatens security and can complicate vulnerability assessments and the implementation of recommendations that arise form them.. Sometime security managers or other security personnel become (or are forced to become) so focused on satisfying auditors, regulations, and formal requirements that they lose sight of real-world security issues (Johnston, 2001a). Compliance mode is difficult to avoid in large organizations and bureaucracies, in old established operations, and for security programs that do not encourage security personnel to be flexible, creative, or proactive.

Vulnerability assessments, like any kind of security analysis, suffer from ambiguities associated with the choice of metrics and with the difficulty of conducting a cost/benefit analysis. In the security field, success is defined as nothing happening. That is a vary bizarre metric, and it makes tradeoffs between costs and benefits difficult to rigorously analyze.

A very common problem in security is that security managers and planners often have a very different idea of what is happening in a security program than what is really going on (Johnston, 1997c). Vulnerability assessors should ideally try to analyze the true security program, not the mental image of the program that exists at high levels in the organization. Doing this, however, can create a lot of resistance from high-level security managers and planners.

Many real-world attacks on security devices or systems rely on false alarming, fault analysis, or "watch and pounce" methods (defined below). Yet because these are anomalous, rare events, they can be quite difficult for vulnerability assessors to observe, model, predict, or replicate. It can also be difficult to sufficiently control related parameters.. The classic example of a *false alarm* attack is for burglars to shake the windows outside a bank building one night to set off the alarm. When they first do this, the police arrive in a hurry to check out the alarm. The next night, and for each of the next 3 nights, the burglars generate the same false alarm. After 5 nights in a row, because of all the false alarms, the police either are slow to arrive, fail to arrive entirely, or the alarm system has been turned off because it has become a nuisance. That is when it is safe for the burglars to actually enter the bank in order to steal money. *Fault analysis* is a method used by an adversary to learn about a security device or system (especially a complex one) and its vulnerabilities by studying how the device or system behaves when exposed to unusual conditions or probes. *Watch and pounce* attacks involve the adversary passively waiting and observing until security personnel make a mistake, then leaping into action to exploit that mistake. In general, it can be very difficult to control experimental parameters for rare events so that meaningful results can be obtained.

It is generally quite difficult to obtain realism when testing or demonstrating vulnerabilities with tags, seals, or security programs. This is particularly true inside high security facilities. One major reason is that for critical applications, such as guarding nuclear weapons, highly realistic tests may be too risky or difficult to arrange inside the facility. Indeed, one of the best times for adversaries to attack a facility is when security personnel think a drill is underway. Another factor that limits realism for tests inside a high security facility—but a factor we welcome—is the need to avoid putting vulnerability testers at risk of injury or death. Real adversaries may not feel so constrained. It is also usually quite difficult to arrange realistic tests inside a large secure facility without alerting security personnel or security committees that such tests are underway. Advance notice can distort the experiment.

There are problems with achieving realism even for tests or demonstrations on seals outside the facility in which they are used. For one thing, security managers often send their best seal inspectors to participate in experiments, rather than their average or mediocre inspectors. In the tests and demonstrations, the inspectors are unavoidably on high alert and often don't use the same seal use protocols they routinely employ. Obviously this can skew results.

A particularly interesting challenge in trying to maintain realism has to do with the artificial paranoia we typically see in experiments for Type 2 vs. Type 1 errors (Johnston, et al, 2001b). In a Type 2 (false accept) error, the seal inspector fails to detect that a seal

has been attacked. In a Type 1 (false reject) error, the inspector incorrectly believes a seal has been attacked when it really has not. Usually there is some kind of inherent tradeoff between Type 1 and Type 2 errors, such as shown schematically in figure 1.

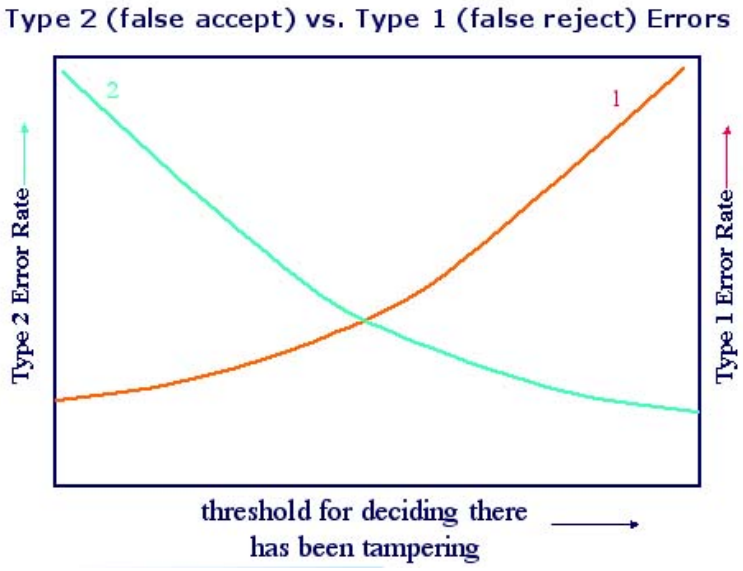

Figure 1 - Low rates of Type 2 errors typically come at the cost of relatively high rates for Type 1 errors, and vice versa

For most tamper detection applications, Type 1 errors are approximately as undesirable as Type 2 errors. Seal inspectors usually recognize that accusing a seal of having been attacked, when it has not, can have serious repercussions. (There are, however, applications where Type 1 errors are much less serious than Type 2 errors. If, for example, the items being monitored for tampering are relatively inexpensive and can be readily discarded if suspicious, or if extensive postmortem forensics are available to reliably check for tampering, high rates of Type 1 errors may be entirely acceptable.)

Another common problem is that sponsors of vulnerability assessments are rarely willing to commit the time and money necessary to conduct thorough and rigorous blind or double blind tests of whether seal inspectors can detect the attacks. Sponsors are usually content with hearing a description of the attacks, seeing them demonstrated, or examining attacked seals—followed by a discussion of possible countermeasures (Johnston, et al., 2002).

A final complication in designing and executing vulnerability assessments has to do with the fact that tags and seals are rarely used in isolation. They are often employed in conjunction with a number of other security devices, personnel, and various nested security layers. Seals, for example, may be used inside a facility that is surrounded by security fences and protected by a guard force and intrusion detectors. The interactions and coupled vulnerabilities among these various security layers can be extraordinarily difficult to analyze.

Reporting Results

One of the major complications in reporting the results of vulnerability assessments is that a defeat of a tag, seal, or tamper detection program is a matter of degree and of probability (Johnston, et al., 2002). A crude attack will not necessarily be detected with 100% probability, nor will a subtle attack always be missed.

We have attempted to deal at least partially with this problem by developing what we call the "Los Alamos Defeat Categorization Scheme" (Johnston, et al., 2002; Johnston, 1997b). Under this scheme, a defeat is classified as being of type 1, 2, 3, or 4 depending on whether it fools the seal inspector when he/she: follows the nominal, usual, or recommended inspection procedures (type 1 defeat), does an unusually careful visual inspection of the exterior of the seal (type 2 defeat), opens the seal and does a careful visual inspection of the seal interior and exterior (type 3 defeat), or uses state-of-the-art forensics techniques to analyze the seal and look for signs of tampering (type 4 defeat). A type 4 defeat is problematic in that it is not possible to *prove* there is no technology capable of detecting the attack. We have nevertheless categorized about 15% of the 289 seal defeats we have demonstrated on 198 different seal designs as type 4 because we are unable to identify any method or technology that could be used to spot the attack (Johnston, et al., 2002).

We have attempted to identify the attributes of an effective vulnerability assessment of a tag or seal, and how to most effectively present the findings and recommendations (Johnston, 1997b). Some of the information that we consider essential in reporting on vulnerability assessments include the following:

- Who did the vulnerability assessment and what is their background and qualifications?
- Do the vulnerability assessors have any potential conflict of interest?
- How many attacks were devised, partially demonstrated, fully demonstrated, and practiced to perfection?
- What was the time and cost to devise, develop, and practice each attack?
- What is the time and cost to execute each attack?
- How much off-site preparation time is needed to execute each attack?
- What inside information, if any, was used for each attack?
- Is each attack high-tech or low-tech in terms of methods, tools, and attack personnel?
- What are the size, weight, and nature of the attack tools and materials for each attack?
- What are the countermeasures and recommendations that arise from the study?
- Vulnerability assessors should try to provide samples of attacked seals, as well as in-person or video demonstrations of the attacks.

• Vulnerability assessors should also provide a sanitized (unclassified) summary of the vulnerability assessment that is devoid of sensitive details. This permits others to judge the thoroughness of the assessment without giving away vulnerability information that might assist adversaries.

Conclusion

Conducting vulnerability assessments of tags, seals, and tamper detection programs is a complex and challenging process. Issues of how to design vulnerability experiments, analyze the results, reach rigorous conclusions, and present findings in a statistically meaningful way are largely unresolved. Addressing these issues is important because of the continuing need for effective tamper detection.

References

Hutchings, Edward (Editor), Leighton, Ralph, Feynman, Richard Phillips, and Hibbs, Albert, 1985, *'Surely You are Joking, Mr. Feynman': Adventures of a Curious Character* (Batam, New York, 1985), pp. 119-137.

Johnston, Roger G. and Garcia, Anthony R.E., 1977a, "Vulnerability Assessment of Security Seals," Journal of Security Administration, Vol. 20, No. 1, June 1997, pp. 15-27, http://lib-www.lanl.gov/la-pubs/00418796.pdf.

Johnston, Roger G., 1997b, "Effective Vulnerability Assessment of Tamper-Indicating Seals," Journal of Testing and Evaluation, Vol. 25, July 1997, pp. 451-455, http://lib-www.lanl.gov/la-pubs/00418792.pdf.

Johnston, Roger G., 1997c, "The Real Deal on Seals," Security Management, Vol. 41, September 1997, pp. 93-100, http://lib-www.lanl.gov/la-pubs/00418795.pdf.

Johnston, Roger G., 2001a, "Tamper Detection for Safeguards and Treaty Monitoring: Fantasies, Realities, and Potentials," The Nonproliferation Review, Vol. 8, No. 1, Spring 2001, pp. 102-115, http://lib-www.lanl.gov/la-pubs/00367047.pdf.

Johnston, Roger G., Martinez, Debbie D., and Garcia, Anthony R.E., 2001b, "Were Ancient Seals Secure?," Antiquity, Vol. 75, No. 288, June 2001, pp. 302-303, http://lib-www.lanl.gov/la-pubs/00818331.pdf.

Johnston, Roger G., 2001c, "The 'Town Crier' Approach to Monitoring," Report LAUR-01-3726 (Los Alamos, NM: Los Alamos National Laboratory, July 2001).

Johnston, Roger G., 2001d, "Tamper-Indicating Seals for Nuclear Disarmament and Hazardous Waste Management," Science & Global Security, Vol. 9, No. 3, 2001, pp. 105-107, http://lib-www.lanl.gov/la-pubs/00818333.pdf.

Johnston, Roger G., Garcia, Anthony R.E., and Pacheco, Adam N., 2002, "Efficacy of Tamper-Indicating Devices," The Journal of Homeland Security (in press).

Jones, James L., 1996, "Improving Tag/Seal Technologies: the vulnerability assessment component," Report 95/00599, (Idaho Falls, ID: Idaho National Engineering and Environmental Laboratory, December, 1996).

Naval Facilities Engineering Services Center (NFESC), 1997, "Antipilferage Seal User's Guide" (Port Hueneme, CA: October, 1997), http://locks.nfesc.navy.mil/Security_seals/guides/seal_ug/rev_sealguide.pdf.

Naval Facilities Engineering Services Center (NFESC), 2000, "DoD Training Course on Effective Seal Use" (Port Hueneme, CA: Spring, 2000), http://locks.nfesc.navy.mil/Security_seals/security_seals/sp2086.pdf.

Tyska, Lou (Editor), 1999, Guidelines for Cargo Security & Loss Control (Annapolis, MD: National Cargo Security Council, 1999), pp. 29-38.

Vulnerability Assessment Team (VAT), Argonne National Laboratory, http://www.ne.anl.gov/capabilities/vat/.