HOMELAND SECURITY INSTITUTE

Journal of Homeland Security

The Problem of Errors, DHS, and the 'False Positive' Standard

October 2007

http://www.homelandsecurity.org/newjournal/articles/giermanski_dhs_false_pos.htm

Dr. James Giermanski, Director, Centre for Global Commerce, Belmont Abbey College

Dr. Peter Lodge, Chairman, Department of Sociology, Criminal Justice and Security Studies, Belmont Abbey College

The Problem of Errors in Testing

Errors or mistakes occur in just about any human endeavor, including medicine, the legal system, scientific research, and even areas that concern the Homeland Security Department. An example from medicine will serve to illustrate the range and the types of errors. Men over age 50 may well have a prostate-specific antigen test to screen for prostate cancer. Under normal circumstances, we assume that a low level of antigens indicates the absence of prostate cancer, and high levels alert us to the possibility of cancer (or, at the very least, the increased risk of developing cancer). Typically, a finding of high levels of prostate-specific antigen would lead to additional diagnostic testing. However, two errors might happen in such cases. One is that although the test indicates worrisome levels of antigens, further diagnostic testing finds no evidence of cancer. This is a false positive,[1] which increases anxiety and leads to additional (unnecessary, as it

happens) costs. As irksome, worrisome, and expensive as false positives might be, we typically tolerate them because alternative tests could subsequently establish that we are cancer-free. The other error that can occur is a false negative[2]—that is, a failure to uncover a true high level of antigens and the failure, therefore, to discover what might be an actual case of prostate cancer. Our concerns regarding undiagnosed disease and the accompanying consequences mean that in matters of health we would prefer to deal with the inconvenience of secondary testing so that cases of cancer are not missed. All things being equal, the lower the tolerance for a false positive, the greater the probability of a false negative.[3] Said the opposite way, the more we allow false positive mistakes in detection (such as indications that a bomb was in a container when, in fact, it wasn't), the less likely we are to have a false negative (that is, to find a bomb in a container said to be without one).

The Situation

Before we look at some implications of these types of errors for container security, we need to look briefly at the situation regarding port security. On October 13, 2006, President Bush signed into law the SAFE Port Act of 2006. In it are some important and interesting definitions. Three terms stand out: international supply chain, radiation detection equipment, and container security device. Section 2[4] of the act defines them as follows:

International supply chain: "The term 'international supply chain' means the end-to-end process for shipping goods to or from the United States beginning at the point of origin (including manufacturer, supplier, or vendor) through a point of distribution *to the destination*" (emphasis added).

Radiation detection equipment: "The term 'radiation detection equipment' means *any* technology that is capable of detecting or identifying nuclear and radiological material or nuclear and radiological explosive devices" (emphasis added).

Container security device: "The term 'container security device' means a device, *or system*, designed, at a minimum, to identify positively a container, to detect and *record*

the unauthorized intrusion of a container, and to secure a container against tampering throughout the supply chain. Such a device, or system, shall have a low false alarm rate as determined by the Secretary" (emphasis added).

Of the three, the definition of a container security device may signal both good and bad characteristics. It is good because it is broad enough to acknowledge the existence and application of a system and not merely the use of a device that simply detects breaches into a container throughout its movement via the global supply chain. It is bad because it stated, "Such a device, or system, shall have a low false alarm rate *as determined by the Secretary*" (emphasis added). The weakness is that, given the realities of existing smart container systems, which by their nature are mechanical and are used in environmentally challenging movements on multiple modes of carriage, the false positive threshold as evidenced by existing Homeland Security Department standards may likely be unreasonable and antithetical to security of the supply chain.

First, we need to appreciate the desirable elements that constitute a smart container system. Second, we need to understand the meaning and effect of the false positive threshold and its impact on genuine security for our ports, instead of academic indoor applications of scientific manipulation. Third, we need to ask what should be a reasonable rate of failure, given the state of construction and technological application.

What Constitutes a Smart Container?

It seems there are no standards to define a smart box. In light of this, we think it appropriate to focus on what a smart box should do. If there could be agreement on that, perhaps a definition would be more easily achievable. In our view and in the view of others, a smart container must perform at least seven clearly defined operations:

1. A smart container must be a part of a system approach necessary to coordinate all facets of the supply chain process to ensure visibility and security. That begins at origin. Therefore, the container must be able to record the identity of the person responsible for monitoring the "stuffing" and securing of the container at the foreign point of origin.

2. There should be an electronic capturing of certain trade data that will link to other documentation. Examples would be the container number or booking number. We could even include portions of the Inward Cargo Declaration, Customs Form 1302.
3. Consistent with the requirements of the Customs-Trade Partnership Against Terrorism to conduct a seven-point inspection of the container, a smart box should be able to detect a breach *anywhere* into its body, not just through the doors.
4. The container should be able to report a breach in real time or close to real time with the date, time, and geographic location of the breach.
5. The smart container is one that can give its geographic position throughout the supply chain when queried, or automatically give its position if it is off its designated course of travel in controlled environments.
6. The container must recognize and record the identity of the authorized person opening the container at the destination.
7. Finally, the container should be adaptable to different sensors and be able to communicate with or be adapted to divergent logistic software packages used by shippers and carriers within the supply chain.[5]

The Problem of Near Perfection

If we understand the complexity of the operation of any system that accomplishes those seven *operations*, we must question a decision by DHS to require a 99% false positive fail rate. In a request for information, an information-gathering and planning vehicle used by DHS in support of Customs and Border Protection, Johns Hopkins University's Applied Physics Laboratory (under contract with DHS) sent a letter dated November 8, 2005, to potential vendors. It stated, in part, "The purpose of this request is to gather information to identify and evaluate available state-of-the-art container and trailer tracking devices suitable for in-bond shipments."[6] That statement, alone, poses two serious questions: What does DHS believe is state-of-the-art, and why did it take so long after 9/11 for DHS to realize that Customs and Border Protection had little or no

knowledge of or control over containers coming into the United States and moving throughout the United States under bond?

State-of-the-art assumes not only the latest versions of technology, but also the latest level of reliability. Therefore, part of the evaluation is the requirement to meet quality standards. In both the request for information in 2005 and in requirements issued in 2006, DHS has determined that a 99% false positive threshold should be the standard. The latest manifestation of this is contained in the Transportation Worker Identification Credential program, requiring special ID cards for port workers and those who have routine access to U.S. seaports. As reported in *Washington Technology*, there is concern over the potential inoperability among cards produced by different manufacturers and the "1 percent system error rate inherent in [Federal Information Processing Standard] 201"[7]—which serves as the guideline in developing the port worker identification card. Some industry leaders are objecting and are concerned about failures in cards that would mean costs and delays at our nation's ports. So why has DHS chosen the 99% false positive standard?

It seems that the 1% or 99% false positive rate is directly linked to the National Institute of Standards and Technology and can be traced to Special Publication 800-76-1. When treating the reliability of fingerprints as part of biometric data, the publication reveals:

Authentication performance is quantified in terms of both the false reject rate (FRR) and the false accept rate (FAR).… FRR is the proportion of legitimate cardholders incorrectly denied access; the latter would be the proportion of impostors incorrectly allowed access.[8]

Specifically in Special Publication 800-76-1, we can see in Section 7.5 that FRR values "are less than or equal to 1% at a fixed 1% FAR operating point." Further, with respect to the interoperability of different fingerprint templates and speed of computation, an "FRR less than or equal to 1% at a FAR of 1%" is required.[9] So it seems we have the basis for the false positive standard.

This standard *might* be reasonable when dealing with biometric identification techniques, but then again, how valuable is a 99% false positive threshold to the safety of our nation's ports and to us as individuals? DHS wants a 99% false positive threshold—that is, DHS allows the device to fail once in 100 uses. This criterion may or may not be the right one. It might be that the false positive rate of 1% or less is unrealistic but that the 1% false negative standard is far too lax. If we accept this standard for the border inspection of containers and trailers for weapons of mass destruction (nuclear, biochemical, etc.), any device used to establish the presence of WMD must have no more than a 1% rate of false positives. Inevitably such a rate would reduce the amount of time (and hence money) devoted to a secondary test (that is, physical search) of the container. However, given the potential dire consequences of a false negative in this example, a 1% false negative is too generous—one dirty nuclear device (assuming a port detonation and depending on the port) could disrupt international commerce at that port for months or years. Which is more disruptive—tolerating a higher level of false positives or the increased possibility of a false negative? (The problem of errors here is further complicated by the kinds of technology employed to monitor incoming containers. For instance, we use portal X-ray machines at our ports, and we know that they do not detect shielded enriched uranium. But we use them anyway. We also know that they alert us to bananas, but we use them anyway.[10])

We should develop the mindset of linking the level of acceptable risk to the potential outcomes. For example, airline passengers expect that the planes they fly in have a 100% probability of landing after they take off. To ensure the validity of this criterion, there are rules for taking off and landing, and planes have redundant systems and emergency systems, and we gladly pay for that. In addition, federal aviation rules control whether a plane can fly given the weather conditions.

Similarly, the scientific community uses different confidence levels for different purposes. Therefore, if we are using a smart container to thwart thefts and hijacking of cargo or for supply chain tracking information, we would likely use and be happy with a 95% confidence level. While the 99% false positive threshold is laudable, the requirement of obtaining near-perfection is extremely difficult in the global container

market and, more important, inhibits the development and implementation of new ideas and practices.

All this talk of errors and of unexpected or improbable occurrences reminds us of some the insights offered by Nassim Taleb, the so-called "skeptical empiricist" and professional derivatives trader, a prolific writer on scientific and philosophical issues.[11] We operate within a world of probabilities that in some ways contributes to the illusion of safety.[12] Taleb argued that we tend to underestimate the likelihood of rare event occurring. But rare events happen, and when they *do* occur they are usually far more devastating than we might have imagined possible.

Some Concluding Thoughts

It seems to us that a more reasonable approach would be to act as if our ports and our country were likely candidates for a particular debilitating disease or as if they were passengers on an airplane.

In the latter case, we could use redundant devices and systems that would make the cost of smart containers out of the question. And we still couldn't control the container's use in bad weather. Thus, even if we were willing to pay, we could not have the reliability of the passenger plane. Maybe we could reach, at best, a 90% confidence level.

With respect to being candidates for disease, the United States has the right mix of freedoms like the right mix of genes, and living in our environment of global violence, which, like air quality, may not be always healthful, we become potential targets for sickness (violence) and death. So what should we do? We look around to find a test, an indicator of potential risk, and a drug that can give us the best protection we can find. If our potential disease is cancer and we can find a drug that works only 80% of the time, should we not take it? Yes, it costs money, and yes, it is not a 100% sure thing, and yes, it may be uncomfortable to take. So Customs and Border Protection doesn't like it. It may cost them, inconvenience them, and even slow down port clearance. So does that mean

we don't take it? We certainly *can* inconvenience some Customs and Border Protection employees whose salaries we pay.

Another concept to consider is the "reasonable man theory" commonly used in judicial or quasi-judicial proceedings. What would a reasonable man believe or do? Said another way, it's the 75% level of probability, or "probable cause." The grand jury system is based on it, and people's lives are affected by it, and the nation accepts it. Would it not be "reasonable" to have a minimum 75% standard of reliability instead of 99%?

We believe that we should institute an allowable failure rate that still gives us the best protection possible. We are dealing with creative, dangerous persons against whom we must have the best defense attainable, not the best defense possible. We cannot afford to have *no* defense because of a bureaucratic standard that simply cannot be met with our current and likely future technology or cannot be developed because of unacceptable costs of added redundancies for sensors and communications. Even with redundant subsystems, containers are victims of our climates, temperatures, rough handling, rough roads, and rough seas. Perfect mechanical and electronic functioning of container security systems capable of all that is expected of them is unattainable at this time. So do we have the luxury of waiting for perfection? I think we all know the answer. We hope that DHS does too.

References

Click on an end note number to return to the article.

1. False positives are also known as Type I errors and false rejection decision errors—the latter because they involve the incorrect rejection of the null hypothesis. To non-statisticians, a "false positive" is usually the preferred name because it is more understandable in words commonly used. A false positive occurs when we decide that something is present in a sample but our decision turns out to be incorrect.

2. False negatives are also known as Type II errors and false acceptance errors. To non-statisticians, “false negative” is usually the preferred name. A false negative occurs when we decide that something is not present in a sample that has been analyzed when, in fact, it is present at detectable concentrations.

3. The relationship can be much more complex, depending on whether there is one distribution for both types of error or two (one for each kind). In the case of medical testing, there may be one distribution that deals with the competence of the physician and another that describes the competency of a particular laboratory.

4. H.R.4954, the SAFE Port Act (enrolled as agreed to or passed by both House and Senate).

5. James Giermanski, “Boxing Clever,” *Cargo Security International*, vol. 4, no. 1, Feb./March 2006, p. 44.

6. An in-bond shipment is “an import or export shipment which has not been cleared by Customs and is transported, stored, or handled with security to the government provided by indemnity bonds”— TeachMeFinance.com.

7. Alice Lipowicz, “More Woes for TWIC,” *Washington Technology*, vol. 21, no. 23, Nov. 27, 2006, p. 1.

8. Charles Wilson, Patrick Grother, and Ramaswamy Chandramouli, “Biometric Data Specification for Personal Identity Verification,” National Institute of Standards and Technology, Special Publication 800-76-1, January 2007.

9. “Biometric Data Specification for Personal Identity Verification,” Section 8.9.

10. James Giermanski, “No More Excuses or Delays,” *American Shipper*, October 2006, p. 2.

11. For a list of his publications since 2004, see Nassim Nicholas Taleb’s Home Page.

12. We may know that the probability of a single engine “o” ring or seal failing is 1 in 100 journeys, but when the engine has four of those seals, then the probability of failure is 1 in 25 journeys, and it may not just be the engine that fails—the consequences could be substantial loss of life.