Risk Analysis, Vol. 26, No. 5, 2006 DOI: 10.1111/j.1539-6924.2006.00817.x

Preventing the Importation of Illicit Nuclear Materials in Shipping Containers

Lawrence M. Wein,[1]* Alex H. Wilkins,[2] Manas Baveja,[2] and Stephen E. Flynn[3]

We develop a mathematical model to find the optimal inspection strategy for detecting a nuclear weapon (or nuclear material to make a weapon) from being smuggled into the United States in a shipping container, subject to constraints of port congestion and an overall budget. We consider an 11-layer security system consisting of shipper certification, container seals, and a targeting software system, followed by passive (neutron and gamma), active (gamma radiography), and manual testing at overseas and domestic ports. Currently implemented policies achieve a low detection probability, and improved security requires passive and active testing of trusted containers and manually opening containers that cannot be penetrated by radiography. The annual cost of achieving a high detection probability of a plutonium weapon using existing equipment in traditional ways is roughly several billion dollars if testing is done domestically, and is approximately five times higher if testing is performed overseas. Our results suggest that employing high-energy x-ray radiography and elongating the passive neutron tests at overseas ports may provide significant cost savings, and several developing technologies, radiation sensors inside containers and tamper-resistant electronic seals, should be pursued aggressively. Further effort is critically needed to develop a practical neutron interrogation scheme that reliably detects moderately shielded, highly enriched uranium.

KEY WORDS: Game theory; nuclear weapons; port security; queueing theory

1. INTRODUCTION

The detonation of a nuclear weapon on U.S. soil is the most feared type of terrorist attack. Standardized shipping containers, which transport over 95% of U.S. imports and exports by tonnage, are highly vulnerable vehicles for delivering nuclear and radiological weapons.(1) The cost of an exploded bomb at a major U.S. shipping port has been estimated to be a trillion dollars,(2) although terrorists may prefer to maximize the human toll by detonating a smuggled weapon in a city center rather than at a port. Several technologies can be used to detect a nuclear weapon and a variety of newer technologies are undergoing rapid development. Moreover, the hourly waiting cost of a container ship arriving at its U.S. port of debarkation is tens of thousands of dollars. Hence, these technologies must be deployed in a way that does not slow down world trade.

Against this backdrop, the U.S. government needs to identify a testing strategy that specifies which containers to test, how to test them (which includes the equipment used and the threshold levels that dictate pass/fail results), where to test them (at the overseas port of embarkation, at the domestic port of debarkation, or both), and how many resources (people and equipment) are required to guarantee, with a high probability, that containers move through the testing process sufficiently fast. As with most homeland

[1] Graduate School of Business, Stanford University, Stanford, CA.
[2] Scientific Computing and Computational Mathematics Program, Stanford University, Stanford, CA.
[3] Council on Foreign Relations, New York.
* Address correspondence to Lawrence M. Wein, Graduate School of Business, Stanford University, Stanford, CA 94305-5015; tel: 650-326-0692; fax: 650-725-0468; lwein@stanford.edu.

security issues, this problem needs to be assessed not only from the U.S. government's viewpoint, but also from the perspective of the terrorists. Hence, our analysis also allows the terrorists to manipulate the security system designed by the U.S. government, and in particular allows them to shield their nuclear weapon.

The multi-agent aspect of the problem (the U.S. government wants to maximize the probability of detection, while the terrorists want to minimize it) leads us to use a game-theoretic approach.[3] The large number of decisions to be taken (eight by the U.S. government, three by the terrorists) and the multi-attribute nature of the problem (costs, waiting times, and detection probabilities need to be traded off against one another) precludes us from simply identifying and evaluating various common-sense policies. Indeed, the resulting game theory formulation is a complex optimization problem that we cannot solve analytically, leading us to resort to a computational approach. We consider a Stackelberg game,[3] where the U.S. government is the leader and moves first (i.e., designs a testing strategy) and the terrorists are the follower and move second (try to manipulate the testing system). This optimization problem chooses a U.S. government testing strategy that maximizes the minimum probability (this minimization is over the various terrorist decisions) of detecting a shielded nuclear weapon in an imported container, subject to constraints on the amount of congestion, or queueing, at the ports of embarkation and debarkation, on the strategy's total cost, and on the amount of terrorist shielding.

Because of the rapid pace of technological development in this area, we split our analysis into two parts. First, we restrict ourselves to existing detection technologies that are in widespread use. We then consider alternative modes of testing, which include existing technology used in a nontraditional manner and developing technologies that may become available within the next several years. In this way, we attempt to address the near-term problem of how to deploy existing technology in addition to assessing the promise of various developing technologies.

The remainder of the article is organized as follows. Section 2 introduces the 11-layer security system. To maintain computational tractability, we restrict ourselves to the six families of testing strategies defined in Section 3. Section 4 describes the weapons, which are former Soviet nuclear warheads containing either plutonium or uranium. The mathematical models for passive (neutron and gamma) and active (gamma radiography) tests are formulated in Section 5, and the queueing models at the overseas and domestic ports are described in Section 6. The costs are specified in Section 7. Our base-case results, which include an evaluation of the current U.S. policy, is given in Section 8. The alternative modes of testing are described and evaluated in Section 9 and the impact of terrorist shielding is investigated in Section 10. Concluding remarks are provided in Section 11.

The detailed mathematical formulation and other supporting material appear in Appendices A–G, which are maintained by the first author at http://faculty-gsb.stanford.edu/wein/personal/container.pdf.

2. THE 11-LAYER SECURITY SYSTEM

Similar to Reference 4, our analysis incorporates 11 layers of protection, which are depicted in Fig. 1. The first three layers attempt to detect an attack before the container enters the overseas port of embarkation. The first layer is a voluntary self-certification system, as dictated by the Customs-Trade Partnership Against Terrorism (C-TPAT) Program, which allows companies satisfying certain security standards to gain the status of a certified shipper. We call this self-certification because it is essentially run on an honor system: resources are not yet in place to externally monitor and validate companies' security efforts (see Reference 5 for more details). Consequently, U.S. Customs and Border Protection views C-TPAT as an incremental means to improve security.[5] We assign a detection probability of $d_c = 0.2$ if a terrorist attempts to infiltrate the container of a certified shipper. The two most likely approaches are for individual terrorists to become truck drivers or for a terrorist cell to set up its own shipping company, which requires at least two years of problem-free shipping before certification. Hence, we are assuming that an attempt is successful with probability $1 - d_c$, either by a terrorist who got a job as a driver of either short-haul or long-haul trucks without arousing suspicion, or by a sleeper cell that successfully started its own certified shipping company. The annual turnover rate for short-haul truck drivers is as high as 300%, causing most trucking companies to forgo security checks for drivers, which can take up to three months. In our model, the terrorists decide whether to insert the weapon into the container of a certified shipper or an uncertified shipper.

The second layer of security is each container's mechanical seal, which allows remote verification of

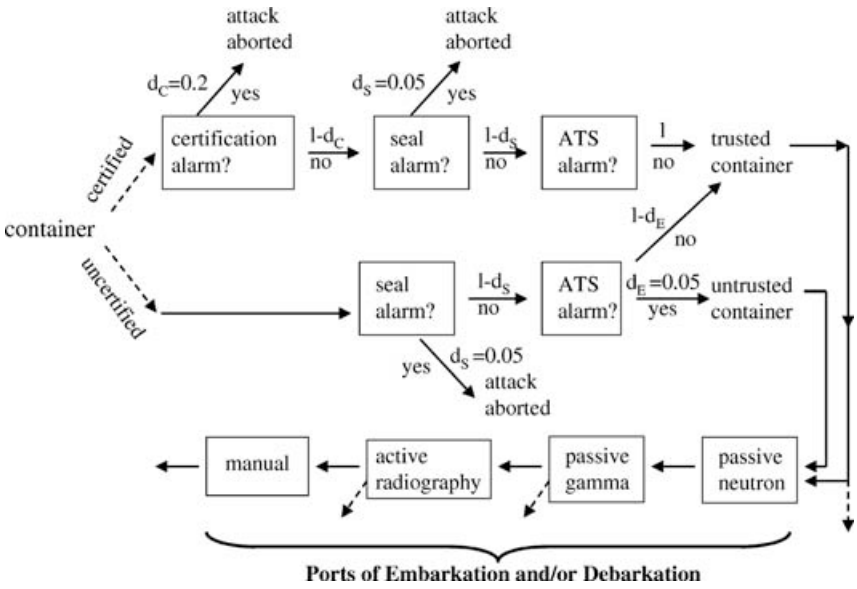

Fig. 1. The 11 layers of security described in Section 2.

serial number but requires manual validation that the seal is intact. This layer is assumed to set off an alarm (e.g., a customs inspector or longshoreman notices that the seal has been tampered with) with probability $d_s = 0.05$ if a terrorist attempts to put a nuclear device into a container, and does not generate any false positive alarms. While a terrorist short-haul truck driver would need to go through this security layer, a terrorist cell setting up a certified shipping company would not. Nonetheless, because the value of d_s is so small, imposing this security layer on a terrorist shipping company does not affect our qualitative conclusions. Although seal technology is in flux, overcoming a mechanical seal at this point in time only requires several minutes with primitive tools,[6] but as the technology improves this may become more difficult; we consider electronic seals in Section 9. We assume that the attack is aborted if one of the first two layers of security is successful, which is reasonable if the weapon is confiscated or if a terrorist cell has to begin the two-year certification process all over again.

The third layer involves an inspection of the documentation that accompanies a container via the U.S. Customs' Automated Targeting System (ATS), an expert system that identifies suspicious containers at the port of embarkation based on their manifest and customs entry document, as well as whether the shipper is self-certified via the C-TPAT Program. All containers shipped from an overseas port of embarkation to a U.S. port of debarkation are deemed to be either *trusted* or *untrusted* by the ATS. Because ATS currently flags 2–3% of containers, we assume that 97% of weaponless U.S.-bound containers are deemed trusted, meaning that they have not triggered the ATS. If a nuclear device is transported on a container of an uncertified shipper then we assume that this software system successfully targets the container, and deems it untrusted, with probability $d_E = 0.05$; however, ATS is assumed to be unable to successfully flag a nuclear device in a container of a certified shipper. This low value of d_E partially reflects the inherent unreliability of the manifest data: shippers may revise their manifests up to 60 days after the container arrives at the U.S. port (see Reference 7 for a description of this and other shortcomings of the ATS). Rather than assuming that d_E is a function of the shielding thickness via the ATS detecting an unexpectedly heavy container, we impose a weight limit on the shielding of 13k lbs, which is 20% of the container weight limit. In contrast to the first two security layers, if a container from an uncertified shipper sets off the ATS alarm, this simply means that the container proceeds through the testing process as an untrusted container, which may affect how aggressively it is tested (see Section 3).

The first three security layers do not incorporate any detailed modeling, and the reader may wonder why they are included in the analysis. The main reason for inclusion is so that we can assess the effectiveness of the currently implemented system, which (as will become clear later) depends critically on these first three layers. Moreover, all three layers are in technological flux, in that their detection probabilities could increase significantly over the coming years: the

U.S. government could transform the C-TPAT program from a self-certification program into a certification program, the mechanical seals could be replaced by electronic seals (see Section 9), and the ATS could disallow manifests to be revised after the container arrived in the United States. By incorporating these components, our model can assess the impact of these improvements. Although the values of the detection probabilities of the first three layers (d_c, d_s, d_E) were chosen somewhat arbitrarily, we chose them conservatively, that is, we erred on the side of overestimating these quantities. Even these conservative estimates are still quite small, and choosing smaller values would not affect our qualitative conclusions.

The last eight layers of protection involve four layers at each port; while the two passive tests are typically performed by one piece of equipment, we view them as two separate layers. These two passive tests measure the emissions of neutrons and gamma rays, respectively, as the container passes through a portal monitor. Because these emissions may be veiled by terrorist shielding, the third layer is active gamma radiography, which emits gamma rays through one side of the container and measures how many of these rays come out the other end, typically allowing it to see dense material, such as the shielding. The final layer at each port is manual testing, where a team of five people open up the container and examine its contents. The precise routing of containers (i.e., the solid vs. dashed lines in Fig. 1) through the last eight security layers depends on the testing strategies in use, which are described in the next section.

3. TESTING STRATEGIES

This section describes the testing strategies at the two ports. We do not consider the possibility of testing at the locations where containers are originally sealed because of the prohibitive cost of reliably enforcing the security of the containers while they are transported to the port of embarkation. In contrast, we assume that there is sufficient security on the ships and at the ports of embarkation and trans-shipment to deter terrorists from attempting to introduce (dangerous and/or heavy) nuclear or radiological material into a container after it has entered the port of embarkation.

The four testing layers at each port are typically viewed as hierarchical, in the sense that the passive tests are the easiest to administer, the active test takes longer but can detect shielding, and the manual test is by far the most expensive but also the most reliable. Because it would be computationally prohibitive to optimize over all classes of strategies that employ these four tests, we use this hierarchy to restrict ourselves to six classes of strategies. Recall that all containers are deemed trusted or untrusted as they enter the port of embarkation. All testing strategies are denoted by the notation $YZ(a)$, where Y describes the set of containers that might be tested ($Y = A$ for *all* or $Y = U$ for *untrusted*, as described below), Z defines where these strategies are applied ($Z = D$ for port of debarkation, $Z = E$ for port of embarkation, and $Z = B$ for both ports), and a specifies the fraction of containers that pass passive testing but are still actively tested; for example, strategy AE(0.4) uses the "A" strategy described below at the port of embarkation, and actively tests 40% of containers that pass passive testing. Hence, we consider six families of policies indexed by a, where

Strategy A: Passive (neutron and gamma-ray) radiation monitoring of all containers followed by active (gamma radiography) testing of all untrusted containers, of trusted containers failing radiation monitoring, and of a fraction a of trusted containers that pass radiation monitoring.

Strategy U: Trusted containers are not tested. Passive radiation monitoring of all untrusted containers, followed by active testing of untrusted containers failing radiation monitoring, and of a fraction a of untrusted containers that pass radiation monitoring.

If a container fails at least one of the passive tests, then it undergoes subsequent tests until the reason for the failure can be ascertained. We assume 85% of weaponless containers failing passive testing are successfully diagnosed at active testing and exit the testing system, while the remaining 15% are diagnosed at a subsequent manual test (to our knowledge, no data has been collected to estimate this quantity, and our estimate is based on expert judgment). In contrast, containers that pass both the passive tests proceed from active testing (if they undergo this test) to manual testing only if they fail the active test. In this study, we are not explicitly concerned with exactly *how* containers that fail these tests are correctly identified (i.e., as false positives that set off alarms because of measurement errors or high background rates, as legal shipments with radiological material or heavy material, or as containers with weapons), although we will estimate the cost and time involved in this investigative process.

Presumably, this identification can be done with some combination of gamma spectroscopy isotope identifiers, active imaging that identifies material-specific signatures, the container's declared manifest, communication with the shipper, and manual inspection.

Although the U.S. government has not formulated a specific strategy, it is using a variant of strategy U, that is, only containers that are flagged by the ATS undergo any passive or active testing.[7] This may be due to lack of testing resources.

In Section 8, we formulate the Stackelberg optimization problem that allows us to optimize and evaluate the six classes of policies.

4. THE WEAPONS

A terrorist organization can steal, buy, or otherwise acquire either a nuclear weapon or highly enriched uranium or plutonium; in the latter case, it needs to build a gun-type bomb, which is significantly easier for uranium than plutonium.[8] We assume that a terrorist is trying to smuggle into the United States a nuclear device similar to a Soviet nuclear warhead, containing either 4 kg of weapons-grade plutonium or 12 kg of weapons-grade uranium. These weapon models are taken from Fetter *et al.*[9] and are surrounded by a concentric shell of tungsten (Fig. 2). Although it is unlikely that terrorists will obtain a nuclear warhead from the former Soviet Union, these weapon models are good surrogates for shielded weapons-grade plutonium or uranium because they contain tungsten tampers that both shield emissions and improve the weapon's efficiency. Because terrorists can also ship the weapon in parts, in Section 10 we consider various levels of terrorist shielding, which also serves as a surrogate for shipping the weapon in parts. The 4 kg of weapons-grade plutonium is made up of 93.3% ^{239}Pu, 6.0% ^{240}Pu, and minute amounts of other elements.[9] ^{240}Pu is the primary source of emissions, and the weapon emits $S_N = 400$k neutrons/sec, in addition to $S_G = 600$ gamma rays/sec at gamma ray energy of 0.662 MeV; in contrast, a plutonium weapon with a depleted uranium tamper would emit 60k gamma rays/sec.[9]

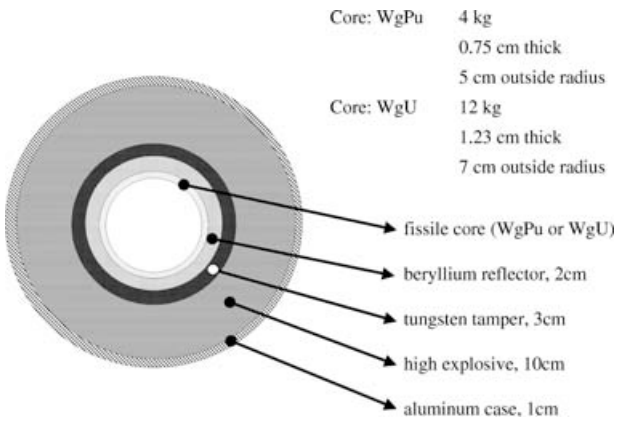

Fig. 2. A depiction of the plutonium and uranium weapons (adapted from Reference 9).

The weapons-grade uranium consists of 93.3% ^{235}U, 5.5% ^{238}U, 1.0% ^{234}U, and 0.2% other elements.[9] Decay products of ^{238}U and ^{237}U are the dominant emissions.[9] Although an unshielded uranium weapon emits predominantly at 185.74 KeV, 3 cm of tungsten causes the dominant emissions to occur at 1.001 MeV via ^{234m}Pa, which is a decay product of ^{238}U. This uranium weapon emits $S_N = 30$ neutrons/sec and $S_G = 30$ gamma rays/sec at the 1.001 MeV level.[9] Some former Soviet nuclear warheads contained uranium from reprocessed reactor fuel (in particular, ^{232}U), which would generate 2.614 MeV emissions that would be easier to detect than the emissions from uncontaminated uranium;[9] terrorists may find it easier to acquire reprocessed fuel than weapons-grade material. The emission rates in Reference 9 were computed using the software package TART.[10]

A modest amount of radiological material (e.g., 10 g of ^{137}Cs) emits almost 10 orders of magnitude more gamma rays than these shielded nuclear devices, as explained in Appendix F. Hence, our analysis focuses on a nuclear device, with the understanding that if an inspection system is reasonably effective at detecting a nuclear source, then it would likely detect a radiological source.

We assume that if terrorists go through the effort of stealing, buying, or fabricating a nuclear weapon, then they will attempt to further shield it with material that lowers the radiation emissions. The tungsten in the weapons in Fetter *et al.*[9] reduces the gamma ray emissions to below the mean background rate, and so there is no apparent need to employ additional tungsten. Lithium hydride is known to be an effective shield for slow neutrons because the lithium-6 in lithium hydride captures the neutrons while producing no gamma rays. To this end, we assume that the weapon is surrounded by r_s cm of lithium hydride. As explained later, more shielding (i.e., larger r_s) leads to more difficult detection by passive monitoring, but perhaps easier detection by active radiography or by noting that the container weight is at odds with the container's manifest. The weight limit of 13k

lbs constrains the shielding thickness r_s to be no more than 84.8 cm (Appendix D).

There are two ways for terrorists to conceal the shielded weapon from radiography: surround it with other metal (e.g., steel) objects, or hide it in a container that has a variety of nonuniform items. We focus on the former here because more legal shipments contain heavy metal than a hodgepodge of items. A standard $40 \times 8 \times 8$ ft container has a weight limit of 65k lbs. A container full of steel (with density 7.8 g/cm^3) would weigh about 20 times this amount. Hence, if the concealed object was in the middle of the container and surrounded by identical steel objects, these objects would need to have a packing fraction (i.e., the fraction of container space consisting of steel) of less than 5% to satisfy the weight limit. Densely-packed heavy-metal shipments typically employ $20 \times 8 \times 8$ ft containers to satisfy the weight limit. Hence, we also consider an identical shielded weapon that is surrounded by identical steel objects inside a 20-ft container, where the packing fraction only needs to be somewhat less than 10%.

In summary, the terrorists have three decisions: the amount of lithium hydride shielding (r_s), whether to put the weapon in a 20-ft vs. 40-ft container, and whether to use a certified or uncertified shipper. Because the terrorists will be opportunistic in obtaining fissile material, we do not view the plutonium versus uranium choice as an explicit decision.

5. DETECTION MODELING

Neutrons and gamma rays emanate from three possible sources: the weapon, the contents of a typical weaponless container, and the background level in the absence of containers. In addition, the measurements from the testing equipment are subject to statistical uncertainty. Passive testing identifies nuclear material by measuring the emissions from a container, and active testing measures the transmission of radiation (at various energies) through a container to discriminate between nuclear or heavy-shielding material and the remaining contents of the container. Appendices A.1–A.3 present mathematical formulas—elaborations of existing formulas[9] to incorporate measurement noise of passive and active testing, containers that drive through portal passive testing, and active testing that allows detection to depend on the size of the weapon—that quantify the amount of emissions measured by passive testing and the fraction of gamma rays detected by active testing, both in the presence and absence of a nuclear device. These models are summarized briefly in this section.

In passive testing, the neutron and gamma emissions from a container as measured by the detector are linear in the testing time and inversely proportional to the square of the distance between the weapon and the detector. As in Fetter *et al.*,[9] we assume that the true background emissions is a normal random variable that is linear in the testing time and has a standard deviation equal to the square root of the mean. We model the emissions from a weaponless container as a log-normal random variable that varies from container to container. Finally, we incorporate measurement noise by assuming that all measurements are normal random variables, where the standard deviation is proportional to the square root of the mean, which is consistent with the observation that measurement errors typically grow with the magnitude of the measurement. The mean of the measurement is the sum of the three sources. Because the mean itself is random due to the variation in the containers and the background, the actual measurement is a mixture of normals. While gamma rays are emitted at a variety of energies, we focus on the 0.662 MeV gamma ray for passive radiation of plutonium and the 1.001 MeV for uranium.

For concreteness, we assume that active testing employs gamma-ray radiography, although x-ray radiography, which is considered in Section 9, is also used in some ports. We assume a portal gamma radiography machine emits gamma rays in a horizontal direction through the container (i.e., through 8 ft of the container, as it passes lengthwise through the portal), and (on the far side of the container) detects gamma rays that have not interacted with the container contents (i.e., have not degraded in energy). These detectors detect emissions at particular energies where there is a large discrepancy between heavy material (fissile material or shielding, such as lead and tungsten) and light elements. This discrimination is achieved by using the mean free paths (at various energy levels) of various elements, and we focus on the photon energy of 1.3 MeV, which is the energy considered in SAIC's most recent VACIS machines.[11] In our mathematical model of active testing, the probability of any particular gamma ray being detected on the far side of the container depends on the thickness and mean free path of each object in the container (i.e., for each object, the probability is $e^{-\mu r}$, where μ^{-1} is the mean free path and r is the thickness). Then we aggregate the number of detected rays over the weapon area, and use the Poisson approximation to

the binomial distribution. In contrast to passive testing, active testing can cause alarms for a variety of reasons that are independent of nuclear or radiological emissions. Consequently, we assume that 5% of actively tested containers set off an alarm regardless of emissions, simply because an unexpected or mysterious object is seen.[12] In addition, we assume that 10% of the containers are difficult to penetrate because of the contents' density. In the model, these dense containers are 20 ft long and filled with steel at a 10% packing fraction, and are a surrogate for some 40-ft containers (e.g., certain agricultural shipments, consolidated shipments filled with a variety of items) that are either too dense to penetrate or too difficult to decipher.[13]

The test threshold levels, one for each of the three tests, are decision variables in our model. A more stringent threshold level increases the detection probability, but also increases the false positive probability, which leads to more downstream testing, thereby increasing costs and queueing. For all three tests, the false positive probabilities and the detection probabilities are derived in Appendices A.4 and A.5, respectively. For each passive test, we choose the threshold level so that the false positive probability is the same for both 20-ft and 40-ft containers. While the threshold levels for the two passive tests are in terms of emission levels, the threshold level in the active test contains a behavioral component that determines whether or not to manually open a container that cannot be penetrated by radiography (i.e., very few gamma rays are detected on the far side of the container). Hence, the weapon can be detected by an active test in one of two ways. First, if the weapon-free portion of the container is too dense then a sufficiently aggressive strategy against impenetrable containers will decide to manually open the container. Alternatively, the weapon is detected if we can distinguish between the weapon's measurement and the weapon-free portion at the 5% significance level.

We assume that the results from each of the three tests are independent, which is justified by the facts that weaponless containers have no shielded nuclear materials (if they did, we would expect these containers to fail passive and active tests) and that passive neutron detection sets off so few alarms (Appendix A.1). However, to the extent that results of the two passive tests are positively correlated, we are making the conservative error of overestimating the frequency of false positives in the testing process.

Extensive controlled and field experiments for passive testing[14] and information on the capabilities of passive and active testing[15,16] are used to estimate the parameter values in the detection models, and allow us to compute detection probability versus false positive probability curves (these false alarms may be due to measurement errors, high background levels, or the legal contents of weaponless containers) for all three tests in terms of the test threshold levels (i.e., the level that defines an alarm), the amount of terrorist shielding, and the container length; Figure 3 displays these six curves for the plutonium weapon. As shown in Figs. 3a and 3b, passive testing is extremely effective if there is less than 10 cm of lithium hydride shielding because a typical weaponless container emits no neutrons. But with more than 20 cm of shielding, the shielded weapon emissions get lost in the background emissions and passive neutron testing is nearly useless. Passive gamma testing, on the other hand, is ineffective even in the absence of lithium hydride shielding (Figs. 3c and 3d) because of the high background level and because contents of some legal containers have higher emissions than the weapon; as noted in Appendix F, however, passive gamma testing is still needed to detect a radioactive dispersal device, which emits orders-of-magnitude more gamma rays than the background rate. Figs. 3e and 3f reveal that gamma radiography can successfully penetrate a 40-ft container filled with steel items at a packing fraction of 5%, but cannot penetrate a steel-filled 20-ft container with a 10% packing fraction. Hence, gamma radiography can achieve a 100% detection probability with a 5% false positive probability for a 40-ft container, and a 100% detection probability and a 15% false positive probability for a 20-ft container.

6. PORT CONGESTION

Congestion is quantified by the steady-state sojourn time distribution, which is the probability distribution of the amount of time a typical container spends in the testing process. The testing process at the ports of embarkation and debarkation are modeled as queueing networks in Appendices B.2 and B.3, respectively, and the parametric-decomposition approach,[17] which is described in Appendix B.1, is used to approximate the steady-state sojourn time distribution for these queueing networks.

More aggressive testing strategies lead to higher congestion in the testing process. The queueing analysis is only concerned with containers that do not contain illicit nuclear shipments, and the false positive probabilities of the various tests (Appendix A.4) dictate the arrival rates to the various queues. At the

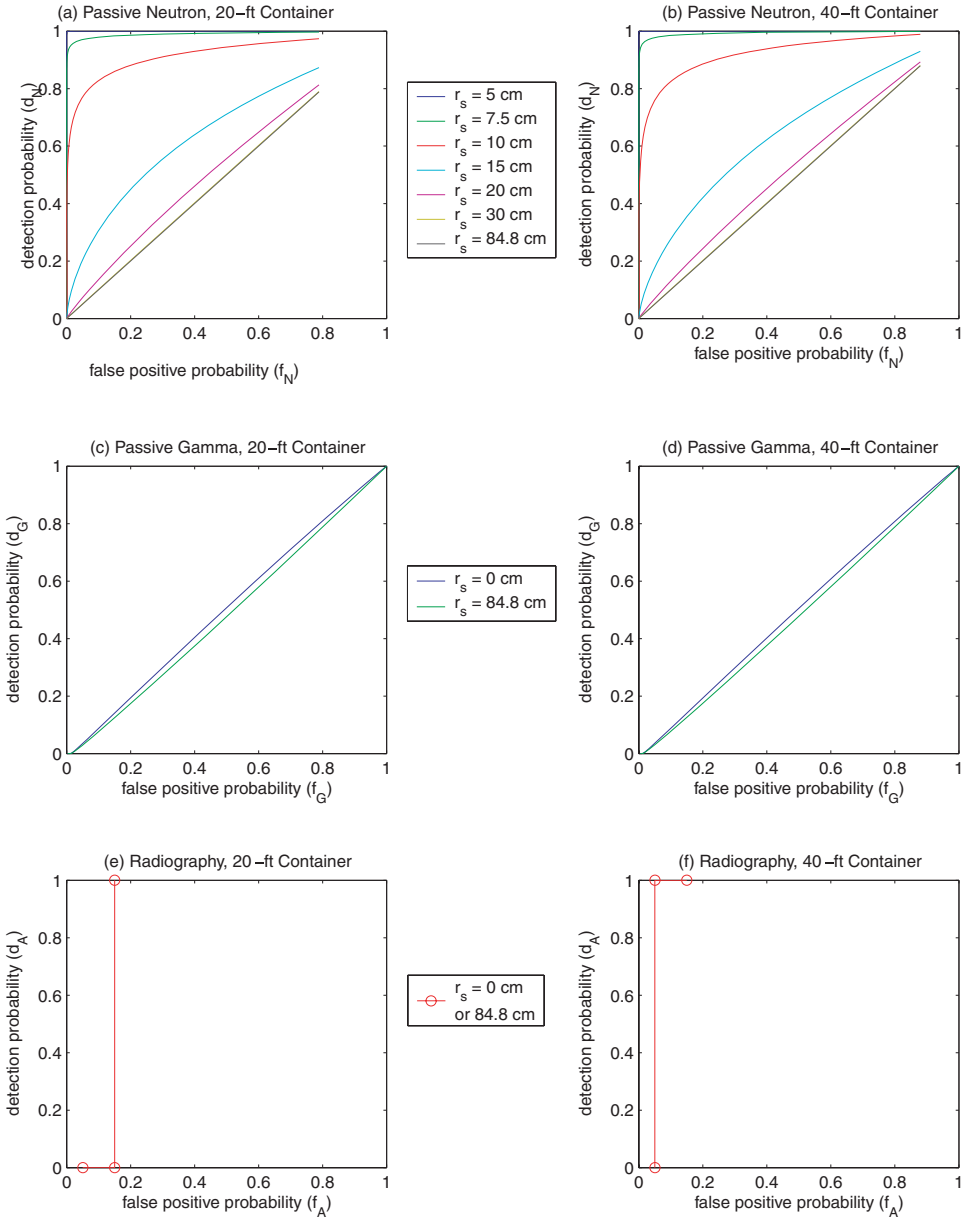

Fig. 3. For the plutonium weapon, detection probability vs. false positive probability as a function of terrorist shielding thickness (r_s), for all three tests (passive neutron, passive gamma, and active) and for both 20-ft and 40-ft containers.

port of embarkation, we are only concerned with U.S.-bound vessels; even Singapore only sails about two U.S.-bound ships per day,[18] and busy U.S. terminals typically unload one ship at a time. Hence, we consider the congestion that occurs at a port due to one ship in isolation, and in Section 7 we scale the costs up to a global basis. We assume that each ship loads or unloads 3,000 containers. A large ship's capacity is about 3,500 40-ft containers, but most ships load and unload containers at several ports, and 3,000 is typical at busier ports such as Hong Kong and Long Beach.

One nonobvious aspect of our queueing models is the determination of the arrival rates of containers. At the port of embarkation, containers arrive according to an appointment system with a 1-hour time window, and typically arrive 6–8 hours before being loaded. We assume four ship cranes can perform 30 moves/hour, so that the 3,000 containers can be loaded onto the

ship in 25 hours. Assuming a maximum waiting time for any container of 8 hours and that 2 hours of work is waiting at the ship crane before it starts loading, so as to avoid ship-crane idleness, we perform a deterministic analysis (Section B.2 and Fig. 1 in the Appendix) to derive the arrival rate of containers to the port of embarkation. The arrival rate of containers to the testing process at the port of debarkation is dictated by the unloading rate from the ship. We assume three ship cranes, each making 30 moves/hour, unload the ship, so that the arrival rate is 90/hour.

Passive testing requires a truck to pass through the portal at about 5 mph, i.e., 5.5 seconds for a 40-ft container. Because it would take about 20 ship cranes to generate containers at this rate, and only three to five cranes are typically used per ship, we can safely disregard any congestion due to passive testing. Consequently, the queueing networks described in Appendices B.2 and B.3 have two stages, one for active testing and one for manual testing. The service times for active testing are assumed to be Erlang of order 2 with a mean of 3 minutes, and the service times at manual testing, where each server represents a 5-person team, are exponential with mean 1 hour. The exponential assumption reflects the fact that the root cause of the upstream testing alarms can often be found without unloading the entire container.

The congestion constraints require that 99% of containers must spend less than 6 hours in the testing process at the port of embarkation, and 95% of containers must spend less than 4 hours in the inspection process at the port of debarkation. Ships have detailed loading plans: each container is destined for a particular three-dimensional location on the ship (containers are stacked approximately 11 high), depending upon its origin, destination, contents, weight, etc. Each container is scheduled to arrive at the overseas port approximately 6 hours before it is due to be loaded. If it is held up in the inspection process then it will be loaded in an overflow area rather than in its planned location. These overflow containers effectively delay the loading and subsequent unloading processes. At domestic ports, containers typically spend 4 hours in the shipyard after getting unloaded, and so both congestion constraints should prevent any significant drop in port efficiency. The higher service level overseas reflects the fact that a container may miss its outgoing boat, whereas it may only delay a truck driver at the domestic port. The number of active testers and the number of manual testing teams at each port are decision variables that enable the congestion constraints to be satisfied.

A key design issue is the location of testing at the overseas and domestic ports. We assume that testing at the port of embarkation is done at the gates leading into the terminal, rather than in the shipyard, for two reasons: the logistics of testing are much simpler, and hence will likely lead to less congestion (e.g., no extra burden on transtainers and utility trucks, which are highly worked now that the 24-hour manifest rule has led to an increase in the number of containers in overseas shipyards), and the impact of a smuggled weapon is less (the container is farther from the berths and ships, and hence is capable of causing less economic—and hopefully human—damage).

The testing logistics at the port of debarkation are more complex because we are forced to take the containers off the boat before testing them. Roughly 60% of imported containers leave the port of debarkation via truck, and 40% by rail. There are several options of where to do passive and active portal monitoring at the port of debarkation. One possibility, as is currently done in several U.S. ports, is to test truckbound containers at the outgoing truck gate and place portal monitors directly on the rail system. However, this option allows a terrorist truck driver to access the container before inspection (and hence detonate the bomb in the most desirable location within the port), and an alarm generated by a testing system on the railroad tracks will delay the entire train. This option also prevents railbound and truckbound containers from being tested on the same equipment, thereby increasing costs.

Railbound containers only incur two crane movements (a bombcart takes it from the ship crane directly to a tophandler, which is a small portable crane that places the container on a train), whereas truckbound containers incur three crane movements (a bombcart takes it from the ship crane and drives it to the shipyard, where it is unloaded by a tophandler; later, a tophandler loads the container onto a utility truck, which moves it from the shipyard to a parking space). So there are two other possibilities for portal monitoring: test the bombcarts, or add two more crane movements to railbound containers (i.e., essentially treating them in the same way as truckbound containers, by dropping them off in the shipyard and later picking them up) and test the utility trucks. If bombcarts are tested, then inspection delays can idle the ship cranes, which need bombcarts to unload the containers onto. Because ship cranes are the main bottleneck and they work faster than active testers, this option would require enough portal monitors to keep pace with the ship unloading process.

A more robust alternative is to add two more crane movements to railbound containers, and to actively test the utility trucks. This option essentially decouples the ship unloading process from the inspection process, thereby allowing the ship crane to work unimpeded. Since passive monitoring does not significantly increase congestion, we assume that passive monitoring is done on bombcarts and active monitoring is performed on utility trucks. This alternative allows the same equipment to be used on truckbound and railbound containers, and forces only the railbound containers that are actively tested to incur two additional crane movements.

7. COSTS

The total annual global cost of a strategy includes the annual salary of labor that operates the equipment or performs manual inspection, plus 0.2 times the purchase cost of the equipment. The 0.2/year factor is meant to account for the 7–10-year lifetime of the equipment, effectively shortened to 5 years by the risk of obsolescence, and maintenance and upgrade costs. In addition, active testing at the port of debarkation incurs capital and labor costs for additional transportation (utility trucks) and container handling (portable cranes). In Appendix C, all of these costs are scaled up from the single shipment of 3,000 containers to the importation of all U.S.-imported containers from all overseas ports in a year. Our cost figures ignore elements such as installation, training, space, and backup equipment (which might be pooled across terminals), and the equipment (e.g., handheld sensors, isotope identifiers) needed to locate and identify the detected radiation source, and hence should be viewed as underestimates.

More specifically, we assume passive testers cost \$80k each and scale with the number of terminals, which is 50 in the United States and 150 overseas, and active testers cost \$100k and scale with the number of ship cranes, which is 300 in the United States and 900 overseas. Each passive tester requires two employees per shift and each active tester requires three employees per shift, where these employees cost \$75k/year, while each 5-person manual testing team costs \$400k/year. Domestic ports run two shifts per day and overseas ports run three shifts per day. Utility trucks cost \$20k and portable cranes cost \$150k.

These cost estimates should be treated with considerable caution. While we could have used a more refined approach and calculated the purchase, operating, maintenance, and labor costs across the lifetime of the equipment using the time value of money, such precision would be misplaced in our view. The equipment costs (i.e., the price paid by the U.S. government and/or terminal operators) are a moving target, and can change drastically over time depending upon such factors as the number of available suppliers and the quantity discounts offered in exchange for worldwide deployment. We believe these factors dwarf the detailed accounting issues (e.g., time value of money, equipment depreciation), and preclude the capability of making a reliable cost estimate in an industry that is in such flux. Nonetheless, our qualitative results are based on the premise that manually testing a container is much more expensive than testing a container using passive or active testing, and this key assumption is unlikely to change in the foreseeable future.

8. BASE-CASE RESULTS

To summarize, for both plutonium and uranium weapons and for each of the six policies in Section 3, the U.S. government chooses the parameter a, the three threshold parameters (s_N for neutron, s_G for gamma, and p_A for active tests) at the appropriate port(s), and the number of active testers (m_A) and manual testing teams (m_M) at the appropriate port(s) to maximize the minimum detection probability, subject to the sojourn time constraints at the ports, a budget constraint, and a shielding weight constraint, where the minimization is over the terrorist shielding thickness r_s, a 20-ft versus 40-ft container, and a certified versus uncertified shipper. If we let $DP[YZ(a)]$ and $K[\mathrm{YZ}(a)]$ denote the detection probability and total annual global cost for strategy $YZ(a)$, let B be the total annual budget, and let T_E and T_D be the mean sojourn time at the ports of embarkation and debarkation (detailed expressions for $DP[YZ(\mathrm{a})]$, $\mathrm{K}[YZ(\mathrm{a})]$, $\mathrm{T_E}$, and T_D in terms of the decision variables and primitive model parameters appear in the Appendix) then the formulation of the Stackelberg game is (see Appendix D for more details):

$$\max_{\{a,s_N,s_G,p_A,m_A,m_M\}} \quad \min_{\substack{r_s \\ \text{20–ft or 40–ft} \\ \text{certified or uncertified}}} \quad DP[YZ(a)] \tag{1}$$

$$\text{subject to} \quad K[\mathrm{YZ}(a)] \le B, \tag{2}$$

$$P(T_E > 6\,\mathrm{hr}) \le 0.01, \tag{3}$$

$$P(T_D > 4\,\mathrm{hr}) \le 0.05, \tag{4}$$

$$r_s \le \begin{cases} 84.8\,\mathrm{cm} & \text{for plutonium,} \\ 82.9\,\mathrm{cm} & \text{for uranium,} \end{cases} \tag{5}$$

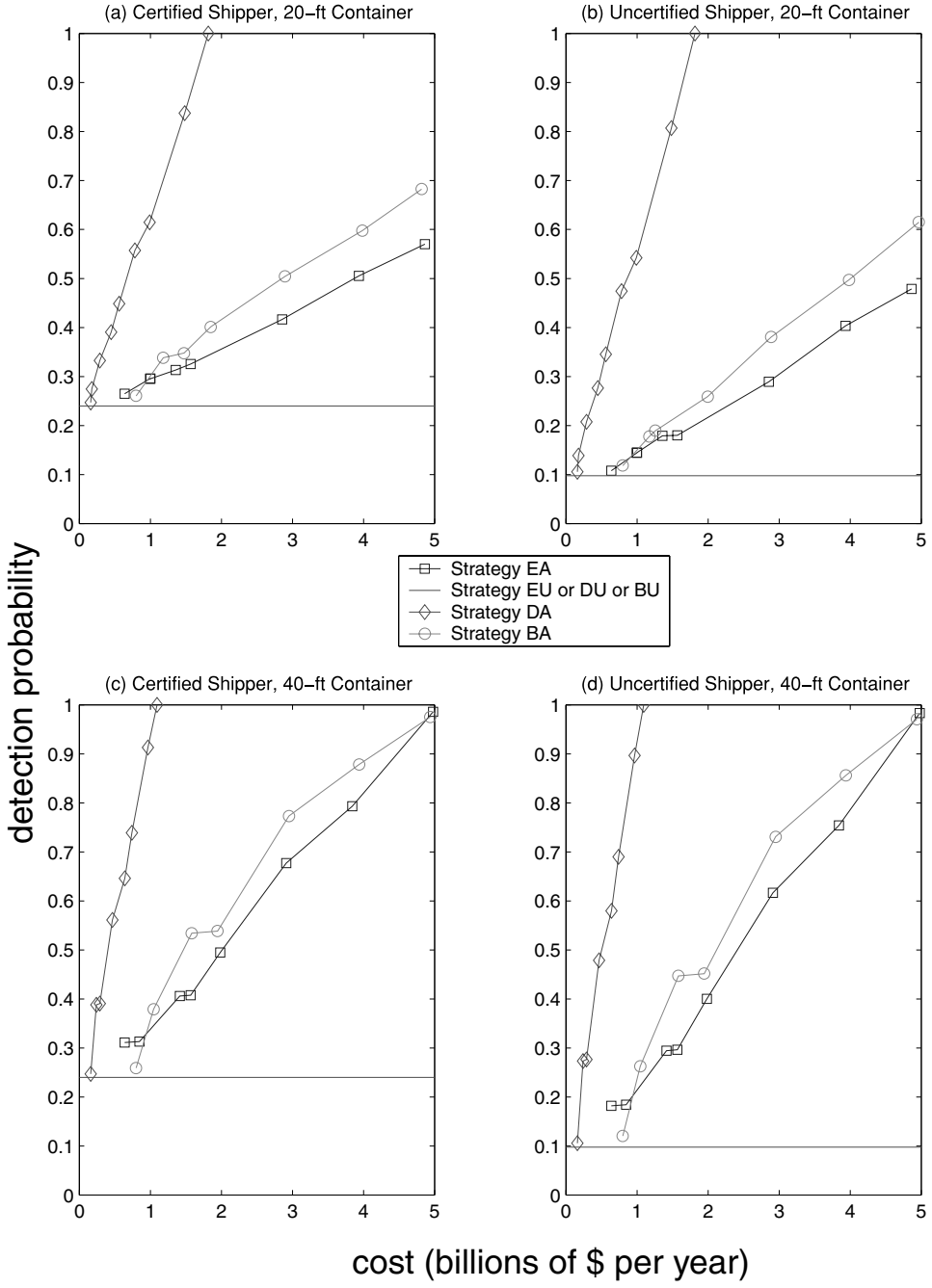

Fig. 4. For a plutonium weapon in a 20-ft or 40-ft container from a certified or uncertified shipper, detection probability vs. cost curves for the six families of policies defined in Section 3.

where Equation (5) is a weight limit that constrains the amount of terrorist shielding.

By varying the available budget B on the right side of Equation (2), we generate in Fig. 4 the detection probability versus cost curves for a plutonium weapon. To understand the impact of two of the terrorist decisions, Fig. 4 considers the four combinations of 20-ft vs. 40-ft container and certified versus uncertified shipper. These curves (the detailed solutions to various points on these curves are in Tables 5–8 of the Appendix) allow for the following observations.

Starting with the terrorist decisions, we note that for a given cost, the detection probability is higher if the container is from a certified shipper than from an uncertified shipper. Similarly, the detection probability is significantly higher for a 40-ft container than a 20-ft container. Hence, the worst-case scenario to protect against is a 20-ft container from an uncertified shipper. In addition, the terrorists use the maximum amount of shielding, 84.8 cm, in all cases, which allows them to evade passive testing.

Turning to the U.S. government decisions, we see that it is significantly less expensive to achieve a given detection level at the port of debarkation than at the port of embarkation. Embarkation detection is more expensive than debarkation detection because there are more overseas terminals and ports shipping to the United States than there are domestic terminals and ports, which limits economies of scale of equipment and labor, and because many overseas ports (e.g, Singapore and Hong Kong) operate three shifts per day rather than two shifts per day (see Equations (44)–(47) in Appendix C). Note that we have also ignored the greater installation costs at overseas ports. Testing at both ports is slightly less costly than testing only at the port of embarkation because (e.g., compare the seventh row and the third-to-last row in Table 5 of the Appendix, which represent an EA policy and a BA policy achieving similar costs) the reduction in overseas manual testing costs more than compensates for the domestic costs incurred.

The three "untrusted" strategies cannot achieve a detection probability higher than 0.0975 or 0.24 if the container is from an uncertified or certified shipper, respectively, regardless of cost or testing location. This is perhaps our most important observation because the current U.S. government policy is an untrusted strategy. The "all" strategies, in contrast, provide increased detection with increasing cost. Radiography can penetrate 40-ft containers and detection is increased in the 40-ft scenarios by increasing the parameter a. Radiography cannot penetrate 20-ft containers, and a high detection probability is achieved by increasing a and opening up the dense containers that undergo active testing. That is, the curves for the "all" strategies in Fig. 4 are generated by increasing the parameter a, which is the fraction of containers passing passive testing that are nonetheless actively tested, from 0 to 1. As the parameter a is increased, more active and manual testers are needed to satisfy the congestion constraints, which leads to exorbitant costs. More specifically, if the terrorists optimally choose a 20-ft container from an uncertified shipper, then a detection probability of 1.0 is achieved by the "all" policy, by either testing at the port of debarkation at a cost of \$1.8B/year, or by testing at the port of embarkation at a cost of \$11.0B/year (not shown in Fig. 4). The cost equations in Appendix C and the solutions in Tables 5–8 in the Appendix enable the detailed cost breakdown among passive, active, and manual testing. Passive testing comprises approximately 2–10% of the total cost and this percentage decreases with increasing detection probability, and manual testing costs are approximately 70% higher (on average) than active testing costs, although the active-to-manual cost ratio in a specific scenario depends on the relative number of servers (m_{EA} and m_{EM}) in Tables 5–8 of the Appendix.

With the maximum amount of lithium hydride shielding, passive testing cannot detect the emissions from either a plutonium or uranium weapon, and active testing can only detect the weapons in a 40-ft container. Consequently, even though the uranium weapon has significantly smaller neutron and gamma ray emissions than the plutonium weapon, the detection probability versus cost curves for the uranium weapon are nearly indistinguishable from those in Fig. 4.

9. ALTERNATIVE MODES OF TESTING

A comprehensive sensitivity analysis is difficult due to the great number of model parameters, the rapid development of new detection technologies, and the uncertainty in the type of illicit weapon used by the terrorists. We perform two types of sensitivity analyses, by considering five other uses of existing technology or hypothetical versions of technology under development in this section, and investigating various levels of terrorist shielding in the next section. Mathematical formulations of four of the five technologies (all except for electronic seals), as well as a discussion of passive monitoring on cranes, are described in Appendix E. As in Section 8, although all results reported in this section are for the plutonium weapon, the results for the uranium weapon are nearly identical.

Several x-ray technologies can achieve deeper penetration and/or better resolution than gamma radiography (e.g., Reference 19). For concreteness, we consider high-energy (9 MeV) x-ray radiography, which can penetrate up to 41 cm of steel,[20] as opposed to 16 cm for gamma radiography. The 9 MeV x-ray equipment can successfully identify the shielded weapon hidden in a 20-ft container of steel items with

a 10% packing fraction. These machines are assumed to cost $1.2 million each and have the same service time characteristics as gamma radiography. Because x-ray radiography is more dangerous to humans (both workers and stowaways) than gamma radiography, we consider a strategy in which all dense containers that are actively tested go to x-ray radiography, and all other containers that are actively tested are handled by gamma radiography. This slightly overstates the benefit of x-ray radiography because inevitably some containers thought to be penetrable turn out not to be, and would be processed by both radiography technologies. Nonetheless, the addition of a x-ray radiography testing layer cuts the cost approximately in half (Fig. 5, which explicitly incorporates detection probability minimization over the 20-ft vs. 40-ft and certified vs. uncertified shipper decisions) by significantly reducing the amount of manual testing that is required.

The majority of active testing time is devoted to analyzing the scan, not producing the scan. Active testing through can be increased by transmitting the scanned images, perhaps remotely, to allow multiple scans to be analyzed simultaneously, thereby decoupling scan production and scan analysis. A three-stage queueing network is used to compute the congestion of this alternative (Appendix E.5). The cost reduction achieved by networked active testing is only about 2–3%.

A variety of electronic seals are being developed and tested, based on radio frequency technology or light sensors, which will generate an alarm if the container has been tampered with. Some of these seals have an intrusion device in the container. Although issues related to power, durability, and effectiveness need to be worked out, we consider a hypothetical $100 tamper-resistant seal that increases the probability of detection to $d_S = 0.95$. If placed in all 12 million containers worldwide, this would achieve a detection probability of 0.96 or 0.9525 for a container from a certified or uncertified shipper, respectively, without any passive, active, or manual testing at either port, at a cost of $200 million/year (Fig. 5). If they are only used on the 12.7% of containers that are untrusted or dense, then the cost is reduced to $25.4 million/year and the detection probabilities are above 0.95 if the weapon is in a 20-ft container or from an uncertified shipper, but is only 0.24 if the weapon is in a 40-ft container from a certified shipper.

Motivated by the small false positive probability of passive neutron testing and the ability to test at the gates at the port of embarkation, we investigate the possibility of improving the detection probability of passive testing by lengthening the testing time at the port of embarkation. Rather than driving through the portal monitor at 5 mph, which takes 5.5 seconds for a 40-ft container, we consider a modification of the EA policy that allows for longer passive neutron

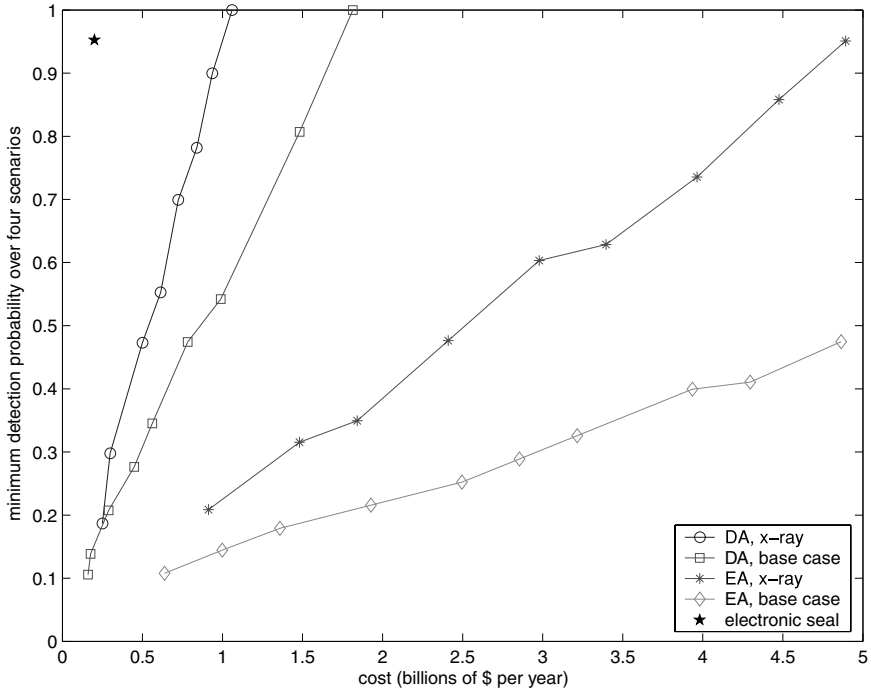

Fig. 5. Detection probability vs. cost curves for several alternative technologies and the plutonium weapon. The DA and EA policies are defined in Section 3.

testing, thereby improving the signal-to-noise ratio. In this case, we explicitly compute the congestion at passive testing (Appendix E.2), and allow additional passive testing equipment to be purchased. In the face of maximum terrorist shielding, this approach offers no significant improvement over the base case.

Lastly, we consider new technology that would replace portal passive monitoring with hypothetical $50 passive monitors (roughly the size and strength of handheld monitors) that are located inside the container and communicate via RFID to an electronic seal or other device, thereby allowing the week-long trip to be exploited for passive (neutron) testing purposes. Radiation sensors require little energy and should last about 10 years with a battery. However, radiation monitors inside of containers have not undergone field tests yet to ascertain their durability in a rugged environment. However, even after seven days of testing in our model, the passive neutron sensor is unable to detect a plutonium weapon with the maximum amount of shielding, and no improvement over the base case is possible.

10. SENSITIVITY TO TERRORIST SHIELDING

It may not be practical for terrorists to obtain 84.8 cm of lithium hydride shielding, and they may use lower-technology shielding such as plywood. Moreover, to the extent that bulkier shielding is less likely to elude the radiography scan analysts, terrorists may prefer to use significantly less than 84.8 cm of shielding. For the base case and two other options—elongated passive neutron testing and passive sensors inside containers—we compute the cost to achieve a detection probability of 0.8 (a level that would likely suffice as a deterrent) as a function of the amount of shielding (Fig. 6). The amount of shielding also serves as a surrogate for smuggling smaller amounts of plutonium or uranium (e.g., splitting a weapon's worth of fissile material across a handful of containers). The cost of elongated passive neutron testing incurs a drastic increase at about 25 cm of lithium hydride shielding (i.e., a shielding factor of 3.4×10^{-4}) in the case of the plutonium weapon. When the amount of shielding is less than this threshold, the cost of obtaining a detection probability of 0.8 is $640 million/year (and is achieved with less than 10 minutes of passive testing per container and very little downstream testing), whereas the annual cost above this shielding threshold is about $9 billion, where detection is achieved by aggressive active testing and some subsequent expensive manual testing. The emissions from the plutonium weapon with 25 cm of lithium hydride shielding is about four times larger than the emissions from the unshielded uranium weapon, and hence no cost fluctuation is observed in Fig. 6 for the uranium weapon. Similarly, a radiation sensor in a container is quite effective against a plutonium weapon with less than 30 cm of lithium hydride shielding (shielding factor = 6.9×10^{-5}) and a uranium weapon with less than 3 cm of shielding (shielding factor = 0.4). In the base case with a plutonium weapon, the cost drops dramatically by shifting from active testing to passive neutron testing when the lithium hydride shielding is less than 10 cm (shielding factor = 0.04). Hence, relative to the base case, elongated passive neutron testing offsets two orders of magnitude of terrorist shielding of a plutonium weapon, and passive sensors in containers offset an additional order of magnitude of shielding.

11. CONCLUDING REMARKS

Our analysis identifies key uncertainties that need to be resolved before a testing strategy can be definitively proposed. These include the fraction of containers that are impenetrable and/or indecipherable by gamma radiography and/or x-ray radiography, the fraction of penetrable containers with a weapon that would be correctly diagnosed by radiography, and the nature of the threat, including the source (uranium vs. plutonium vs. radiological) and the terrorists' shielding capabilities. In this regard, given the difference in detection probability versus cost for the various terrorist decisions (particularly the shielding level), it is imperative that the U.S. government engage in red-teaming exercises (e.g., allowing asymmetric packing of objects surrounding the weapon) to ensure the robustness of any implemented system. Moreover, an estimate of the fraction of legal shipments that emit at the same gamma ray energies as weapons is required to better assess the value of passive gamma testing.

Although our parameter values should be refined with data from ongoing field tests, and the precise nature of a smuggled weapon is largely unknown (at least outside of the intelligence community), our analysis leads to several policy recommendations. First, although there is no currently defined national inspection strategy for imported containers, many containers imported into the United States undergo the UD strategy with $a = 1$; the Container Security Initiative (CSI) enables a limited amount of overseas testing, but again only on untrusted containers. Hence, our model predicts that the likelihood that the current screening system would detect a shielded nuclear

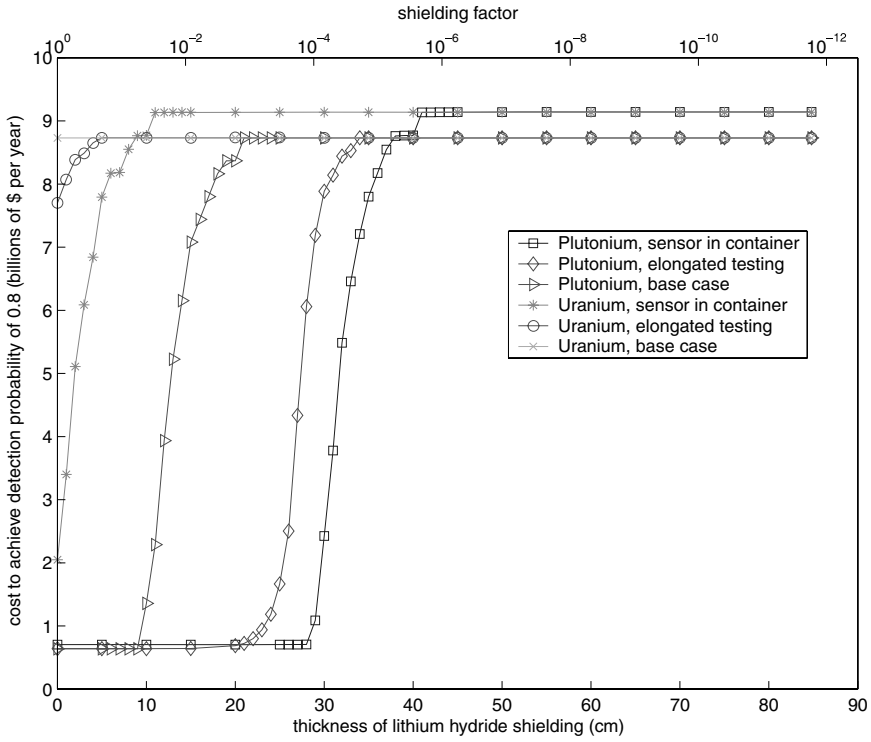

Fig. 6. The annual cost to achieve a detection probability of 0.8 as a function of the thickness of lithium hydride shielding, for overseas elongated passive testing, passive neutron sensors in containers, and the base case, and for a plutonium or uranium weapon.

weapon is quite low (around 10%). While any intelligence information is obviously helpful, the current testing strategy relies almost entirely on the limited nature of the data in the ATS, and in our view is misguided.

If we employ existing technology as it is currently used, two major changes are required to increase our security from the current dismal level to a level that might constitute a verifiable deterrent. First, in the absence of a dramatically improved ATS, trusted containers (i.e., containers that successfully pass the ATS) must be aggressively tested in a layered fashion; although passive neutron testing would detect a lightly shielded plutonium weapon, radiography must also be used to detect other weapons. Second, we must open up containers, trusted or not, that cannot be properly penetrated or deciphered by radiography. These changes require considerable investments, mostly in labor; for example, to achieve at least 90% detection probability in our base case, our rough estimate of the annual cost is $2 billion if testing is at domestic ports and $11 billion if testing is at overseas ports. This domestic versus overseas cost differential needs to be weighed against the increased danger of locating a weapon in the shipyard of a major U.S. port, as opposed to at the gates of an overseas port. Given that terrorists are likely capable of detonating a bomb remotely, it seems that the danger differential outweighs the cost differential, and that most testing should be done overseas. It is possible that this cost differential could be mitigated by using foreign labor to test overseas (maintaining integrity of the testing process would be a significant challenge) and/or to electronically transmit active testing images from overseas ports so that U.S. workers could analyze these images remotely. Regardless of where testing occurs, having a core group of well-trained scan analysts who maintain a database of historical scans (the U.S. government is reluctant to train scan analysts in the threat scenario and these government-owned scans are currently discarded) would likely improve the detection probability of active testing. Pushing testing back beyond the port of embarkation is desirable, to the extent that the port itself is the target of attack. However, this would pose a formidable logistical challenge, possibly requiring a dedicated travel corridor from a centralized testing location to the port.

Our results highlight the fact that the key cost driver when achieving a high detection probability is manual testing labor. Two existing technologies may help to avoid manual testing and hence improve the cost-security tradeoff. First, if the terrorist shielding is not too thick (Fig. 6), elongating passive neutron testing at the gates at the port of embarkation may have the potential to drastically improve this tradeoff, due to the lack of neutron emissions in weaponless

containers, the low equipment and labor cost of passive testing, the simplified logistics of testing at the gates, and the automated nature of passive testing. The automated nature of the passive test enables remote control over the integrity of the testing process and is more robust than active testing, which relies on labor that has been historically trained to detect smuggling of drugs or other contraband rather than nuclear weapons. Our model somewhat naively assumes that if radiography is capable of detecting the heavy shielding surrounding a weapon, then the weapon will be detected; by contrast, when 15 pounds of depleted uranium shielded by a steel pipe with a lead lining was shipped by an ABC News investigative team,[21] the human scan analyst did not detect it even though a subsequent look at the gamma radiography scan clearly showed the object. Given that our passive neutron parameter values are based on ITRAP testing requirements from several years ago,[14] we are likely underestimating the capabilities of passive neutron testing, and hence the elongated passive neutron testing option, which may be capable of a 10-fold reduction in cost (Fig. 6), requires further investigation. The immense cost of manual tests can also be significantly reduced by using 9 MeV x-ray machines on containers that cannot be penetrated by gamma radiography. Also, because x-ray radiography reduces the use of manual testing for those containers that are actively tested, we expect that the cost reduction from x-ray radiography and the cost reductions from either elongated passive testing or passive testing inside a container to be largely independent of each other, for example, if passive testing and x-ray radiography each cut costs by a factor of two relative to the base case, then using both approaches simultaneously should reduce the testing cost by a factor of four.

Turning to developing technologies, putting neutron radiation monitors inside containers has the potential (if issues related to durability, power, tampering, effectiveness, and real-time alarm capabilities can be resolved) to increase detection probability of less than maximally shielded plutonium and moderately shielded uranium weapons in a less expensive manner than aggressive active and manual testing. Although our model assumed that detection would take place at the port of debarkation after a seven-day trip, these sensors might also detect some weapons (except perhaps those consolidated right near the port of embarkation) at the port of embarkation. Indeed, this option could be synergistically combined with elongated passive testing by requiring only the tiny fraction of containers whose sensors detect some neutron emissions at the port of embarkation to undergo elongated passive neutron portal testing, with the great majority of containers undergoing traditional passive neutron portal testing (as an additional layer to protect against tampering with the sensor inside the container). Electronic tamper-resistant seals also have the potential to achieve a high detection probability at about $200 million/year, but to the extent that sophisticated terrorists are likely to figure out how to defeat these seals, it would be prudent to use these seals as part of an overall layered solution. Moreover, a tamper-resistant seal that doubles as a radiation sensor would be an attractive option.

Fig. 6 suggests that while a moderately shielded plutonium weapon can be detected with current technology, an uranium weapon cannot; this is consistent with recent congressional testimony.[22] New technologies such as pulsed fast neutron analysis (PFNA), which exposes containers to short bursts of neutrons and analyzes the gamma-ray signatures that are produced, may prove helpful in uranium detection,[23] although it may take years to develop equipment that is sufficiently small, quick, inexpensive, and safe for practical use. Finally, an alternative approach is to add teeth to the C-TPAT program, using a mix of private- and public-sector inspectors to tightly monitor the process beginning at the factory loading docks, in lieu of extensive testing of containers from certified companies. The estimated cost and detection probability of such a system, which would include electronic seals as a central component,[24] could be compared with the options considered in this article.

In the interim, there are no inexpensive or easy solutions, and one or more of the costlier approaches should be implemented. If these options are deemed too expensive to apply to all containers, then they should at least be applied to all dense, indecipherable, or uncertified containers and a fraction of all other containers, so as to coerce terrorists to introduce the weapon into a container where detection by active testing is somewhat easier. A particularly thorny issue is who pays for these testing systems. Unlike active and manual testing, passive testing does not find any nonnuclear or nonradiological contraband, although it would detect illegal shipments of radioactive waste; a machine that simultaneously performs passive and active testing might obviate this concern, in addition to alleviating space requirements at the ports. Although market forces (e.g., uncertified shippers experience higher insurance rates and/or port delays) may play a role, government mandates backed by government funding (or a security tax passed on to shippers and, ultimately, consumers) are required. Also, the

immense costs required to increase detection suggest that more money should be put into nonproliferation (e.g., accelerating the Nunn-Lugar program, which will spend up to $2 billion/year over the next 10 years to secure nuclear material from the former Soviet Union[25]) and prevention (e.g., a weak link in supply chain security can be strengthened by requiring security checks for overseas short-haul truck drivers, which would need to be performed much more rapidly than 3 months). It is also worth noting that the estimated $10 billion/year required to secure ports is comparable to the current annual investment for ballistic missile defense,[26] particularly in light of the shift in the nature of the threat from adversarial nations to terrorists.

Finally, a solution to this problem requires two types of estimates. First, we need to assess the level of detection probability that constitutes a deterrent, which may be as low as the 0.2–0.4 range. Hence, similar to the inspection of air travelers, the U.S. government should randomly inspect—using passive, active, and if need be manual testing—20–40% of trusted containers, in addition to the 2–3% flagged by ATS. Second, we need to estimate the detection probabilities and concomitant detection costs of alternative smuggling modes (e.g., a fishing boat arriving at the shoreline or arriving by port elsewhere in the Americas and crossing into the United States by land, a private plane detonating over U.S. airspace); these higher-level border issues require as input analyses of the type performed here, but are beyond the scope of this study.

ACKNOWLEDGMENTS

L.M.W. gratefully acknowledges valuable discussions with Matt Bunn, Doug Doan, Jeff Florin, Peter Hsi, Joseph Ng, Vic Orphan, Wade Sapp, and especially John Ochs.

REFERENCES

1. Flynn, S. E. (2000). Beyond border control. *Foreign Affairs 79*, 57–68.
2. O'Hanlon, M. E., Orszag, P. R., Daalder, I. H., Destler, I. M., Gunter, D. L., Lindsay, J. M., Litan, R. E., & Steinberg, J. B. (2003). *Protecting the American Homeland*. Washington, DC: Brookings Institution Press.
3. Gibbons, R. (1992). *Game Theory for Applied Economists*, Princeton, NJ: Princeton University Press.
4. May, M. M., Wilkening, D., & Putnam, T. L. (2004). *Journal of Physical Security 1*. Available at http://jps.lanl.gov/vol1.iss1/1-Container_Security.pdf.
5. Hecker, J. Z. (2002). *Current Efforts to Detect Nuclear Materials, New Initiatives, and Challenges*. U.S. General Accounting Office, GAO-03-297T.
6. Johnson, R. G. (1996). LA-UR-96-2365, Los Alamos National Laboratory, Los Alamos, NM.
7. Stana, R. M. (2004). U.S. General Accounting Office Report GAO-04-557T.
8. Bunn, M., Wier, A., & Holdren, J. P. (2003). *Controlling Nuclear Warheads and Materials: A Report Card and Action Plan*. Cambridge, MA: Project on Managing the Atom, John F. Kennedy School of Government, Harvard University.
9. Fetter, S., Frolov, V. A., Miller, M., Mozley, R., Prilutsky, O. F., Rodionov, S. N., & Sagdeev, R. Z. (1990). *Science & Global Security 1*, 225.
10. Plechaty, E. F., & Kimlinger, J. R. (1976). UCRL-50400, Vol. 14, Lawrence Livermore National Laboratory.
11. Snell, M. P. (1999). Gamma-ray technology: The practical container inspection alternative. *Port Technology International*, *9*, 83–88.
12. Schiesel, S. (2003). Their mission: Intercepting deadly cargo. *New York Times*, *20*, Section E, p. 1.
13. Bjorkholm, P. (2003). Cargo screening: Selection of modality. *Port Technology International*, *17*, 37–39.
14. Beck, P. (2000). Paper OEFZS-G-0005, Austrian Research Centers, Seibersdorf.
15. Richardson, R. D., Verbinski, V. V., & Orphan, V. J. (2002). New cargo inspection and transportation technology application. *Port Technology International*, *15*, 83.
16. Portal Monitors. TSA Systems, http://www.tsasystems.com/products/portals_VM-250A.html, accessed on Oct. 27, 2003.
17. Whitt, W. (1983). The queueing network analyzer. *Bell System Technical Journal*, *62*, 2779–2815.
18. Shipper's Corner. Port Authority of Singapore, http://www.portnet.com/cgi-bin/readcsv1.pl?Country=USA&Submit=Submit, accessed on Oct. 27, 2003.
19. Products and Solutions. American Science and Engineering, Inc., http://www.as-e.com/products_solutions/index.asp, accessed on Oct. 27, 2003.
20. x-ray inspection systems. Smiths Heimann, http://www.heimanncargovision.com/shockedpage.html, accessed on Oct. 27, 2003.
21. Kurtz, H. (2003). ABC ships uranium overseas for story. *Washington Post*, p. A21. Lightly shielded degraded uranium from Jakarta in a container full of furniture was targeted for domestic inspection, and was tested but undetected by radiography. Highly enriched uranium would likely have required more shielding, which could conceivably have been detected by radiography.
22. Huizenga, D. (2005). Detecting nuclear weapons and radiological materials: How effective is available technology? Testimony before the Subcommittee on Prevention of Nuclear and Biological Attacks and the Subcommittee on Emergency Prepardeness, Science and Technology, The House Commmittee on Homeland Security.
23. Neutron technology. Rapiscan Systems, http://www.rapiscansystems.com/neutron.html, accessed on Oct. 27, 2003.
24. Cuneo, E. C. (2003). Safe at sea. *Information Week*, April 7. Accessed at http://www.informationweek.com/story/showArticle.jhtml?articleID=8700375.
25. Luongo, K. N., & Hoehn III, W. E. (2003). Reform and expansion of cooperative threat reduction. *Arms Control Today*, *33*, 11–15.
26. A. L. Tiersky. Missile defense and national security. This appeared in the San Diego Union-Tribune on Jan. 15, 2002. I accessed it at http://www.cfr.org/publication/4289/missile_defense_and_national_security.html.

Appendix

This appendix describes the mathematical model that generated the results reported in the main text. The mathematical models for detection and congestion are given in §A and §B, respectively. The costs are quantified in §C and the strategy optimization and evaluation are defined in §D. Alternative modes of testing are discussed in §E, a radiological source is considered in §F, and supplementary computational results appear in §G.

A Detection Modeling

This section contains mathematical models for three detection processes: passive neutron detection in §A.1, passive gamma-ray detection in §A.2, and radiography in §A.3. All parameter values related to the detection process are given in Table 1. For all three tests, these models are used in §A.4 and §A.5 to compute the false positive probabilities and detection probabilities, respectively.

A.1 Passive Neutron Detection

Emissions emanate from three sources: the weapon, the contents of a typical container that contains no weapons, and the background level in the absence of containers. The monitor is placed $r = 2$ m from the center of the container, which passes through the portal at velocity $v = 2.22$ m/sec (i.e., 5 mph). Hence, if we denote the container length by L, it takes $L/v = 5.5$ sec to monitor a 40-ft (or 12.2-m) container and 2.75 sec to monitor a 20-ft container. After t time units of detection, the true (i.e., ignoring measurement noise) cumulative emissions at the detector due to a stationary source with emission rate S_N that is a distance r from the detector is $\frac{A\epsilon_N S_N t}{4\pi r^2}$ [1], where A is the area of the radiation detector, and ϵ_N is the efficiency for detecting neutrons. Because our source is moving, the true cumulative emissions at the detector will be, for $L = 20$ or 40 ft,

$$\frac{A\epsilon_N S_N}{4\pi} \int_0^{L/v} \frac{dt}{r^2 + (vt - \frac{L}{2})^2} = \frac{\tan^{-1}\left(\frac{L}{2r}\right) A\epsilon_N S_N}{2\pi v r}, \tag{1}$$

if we assume that the weapon is placed in the middle of the container.

We assume that the true cumulative (sea-level terrestrial) background emissions after L/v time units is a normal random variable B_N with mean $A\epsilon_N b_N L/v$ and standard deviation $\sqrt{A\epsilon_N b_N L/v}$ (in units of neutrons) [1], where b_N is the mean neutron background rate.

After L/v time units of detection, we assume that a typical container (containing no weapons) has cumulative emissions at the detector equal to $\frac{A\epsilon_N C_N L}{4\pi r^2 v}$, where C_N is a log-normal random variable (i.e., it varies from container to container) with median e^{c_N} and dispersion factor $e^{\sigma_{cN}}$.

Finally, we incorporate measurement noise by assuming that all measurements are normal random variables (denoted by X_N), where the standard deviation is the unknown factor k_N times the square root of the mean. This relationship holds for a Poisson random variable, and is consistent with the observation that measurement errors typically grow with the magnitude of the measurement. Because the means themselves are random variables due to variation in the container contents and background, X_N will actually be a mixture of normals.

Parameter values for A, ϵ_N and b_N are taken from [1]. To estimate the noise factor k_N, we note that when the detection time is $t = 10$ sec, $r = 2$ meters and $S_N = 20$k neutrons/sec, the false positive probability and false negative probability from extensive controlled experiments (with stationary sources and in the absence of containers and background variation) were 10^{-4} and 10^{-3}, respectively [2]. Substituting these values into

$$P\Big(X_{N1} > s_N\Big) = 10^{-4}, \tag{2}$$

$$P\Big(X_{N2} < s_N\Big) = 10^{-3}, \tag{3}$$

where X_{N1} and X_{N2} are normally distributed with standard deviation equal to k_N times the square root of the mean, and where the means are $A\epsilon_N b_N t$ and $\frac{A\epsilon_N S_N t}{4\pi r^2} + A\epsilon_N b_N t$, respectively, and solving for the two unknowns (the test's threshold level s_N in [2] and the noise factor k_N) yields the value of k_N in Table 1. Finally, we set $c = -\infty$, $\sigma_c = 0$ (i.e., the random variable C_N is always zero) because 163k roadside field tests for container trucks resulted in no alarms [2].

A.2 Passive Gamma-ray Detection

The modeling of passive gamma-ray detection is similar to that of passive neutron detection. We retain the same notation, but use the subscript G in place of N; the values of A, r and v are the same for both types of passive detection. In contrast to passive neutron testing, the remaining parameters for passive gamma testing depend on whether the weapon contains plutonium or uranium. We derive the remaining parameter values for the plutonium weapon first, and then discuss the uranium weapon. While gamma rays are emitted at a variety of energies, we focus on the 0.662-MeV gamma ray in the case of passive radiation, which is the most prominent emission from the plutonium weapon in the main text [1]. The values b_G =1400 gamma rays/m^2·sec and $\epsilon_G = 0.70$ are taken from [1]. The noise factor k_G is estimated from vendor information, stating that the false positive probability is 10^{-3} and 10 grams of ^{239}Pu can be detected with probability 0.5 by a portal monitor with a pillar spacing of 20 ft ($r = 3$ m) at a passage speed of 5 mph in a 20μR/hr background [4]. To find the source term S_G for this experiment, we note that the 0.662 MeV emissions from weapons-grade plutonium are due to a decay product of ^{241}Pu [1] with decay rate 174,000/g·sec [5]. Because 0.44% of weapons-grade plutonium is made up of ^{241}Pu [1], we have $S_G = 0.0044(10)(174,000) = 7656$ gamma rays/sec. To find the background term b_G (we assume the background noise in this experiment is zero), we calculate that 4.53% of the energy from background radiation is at 0.662MeV [6] assuming a 10% energy resolution [1], so that the background radiation should contain 0.0091 μSv/hr at 0.662MeV. Using the conversion factor of 1.0 μSv/hr = 8.94$\times 10^5$ gamma rays/m^2·sec in a NaI detector at 0.662 MeV [7, 8], we find that the background radiation is 7656 gamma rays/m^2·sec at 0.662 MeV. Substituting $S_G = 7656$ gamma rays/sec, $b_G = 7656$ gamma rays/m^2·sec (the two 7656's are coincidental), $r = 3$ m and $L = 40$ ft into

$$P\Big(X_{G1} > s_G\Big) = 10^{-3}, \tag{4}$$

$$P\Big(X_{G2} < s_G\Big) = 0.5, \tag{5}$$

where, for $i = 1, 2$, X_{Gi} is normally distributed with mean $A\epsilon_G b_G L/v$ and $\frac{\tan^{-1}\left(\frac{L}{2r}\right)A\epsilon_G S_G}{2\pi v r} + B_G$, respectively, and standard deviation k_G times the square root of the mean, allows for the determination of the threshold level s_G used by the vendor [4] and the noise factor k_G given in Table 1.

To solve for c_G, we note that field test results [2] state that there were 2256/162,958=0.014 false positives, which was defined as being 15% above background, and 50% of these alarms, or 0.007 of the tests, generated readings greater than 40% of background. Let C_G be a log-normal random variable with median e^{c_G} and dispersion $e^{\sigma_{cN}}$. We model the gamma-ray measurement X_G as a normal random variable with mean μ_G and standard deviation $k_G\sqrt{\mu_G}$, where $\mu_G = \frac{A\epsilon_G C_G L}{4\pi r^2 v} + B_G$, and derive c_G and σ_G by solving (with $b_G = 1400$ gamma rays/m^2·sec, $r = 2$ m and $L = 40$ ft)

$$P\Big(X_G > \frac{1.15A\epsilon_G b_G L}{v}\Big) = 0.014, \tag{6}$$

$$P\Big(X_G > \frac{1.4A\epsilon_G b_G L}{v}\Big) = 0.007. \tag{7}$$

Passive gamma testing of the uranium weapon is done at the 1.001 MeV level, which has $\epsilon_G = 0.57$ and $b_G = 860$ gamma rays/m^2·sec [1]. To compute the source term for equations (4)-(5), we consider 1 kg of ^{235}U [4]. The emissions from weapons-grade uranium are due to a decay product of ^{238}U at 1.001 MeV [1] with decay rate 81/g·sec [5]. Since 5.5% of weapons-grade uranium is ^{238}U [1], we have $S_G = 0.055(1000)(81) = 4455$ gamma rays/sec. To find b_G, we calculate that 7.15% of the energy of background radiation is at 1.001 MeV [6] using a 10% energy resolution [1], so that the background radiation should contain 0.0143 μSv/hr at 1.001 MeV. Using the conversion factor of 1.0 μSv/hr = 6.53×10^5 gamma rays/m^2·sec in a NaI detector at 1.001 MeV [7, 8], we find that $b_G = 9338$ gamma rays/m^2·sec. Using these parameter values, we re-solve (4)-(7) to get $k_G = 0.069$, $e^{c_G} = 1.62$ gamma rays/sec and $e^{\sigma_{cG}} = 43.56$.

A.3 Radiography

Active radiography involves detailed engineering issues related to filtering algorithms, contrast detail and spatial resolution that vary across companies, and developing a mathematical model of this complex process would be a daunting dask. Our objective is to develop a rather simple mathematical model of radiography that allows the detection probability to depend on the size and composition of the weapon and its shielding, and allows the model parameters to be readily estimated from industrial data. In our model, we assume that N_A gamma rays per cm^2 are emitted through one side of a sequence of J solid objects, where object $j = 1, \ldots, J$ is of thickness r_j and has mean free path of gamma rays μ_{jG}^{-1}. The probability of any particular gamma ray being detected on the other side of the sequence of objects is given by g_A, which is approximated by [1]

$$g_A = \prod_{j=1}^{J} e^{-\mu_{jG} r_j}. \tag{8}$$

Hence, if each gamma ray behaves independently, then the observed output X_A from gamma radiography, which is the number of rays detected from a cm^2 of cross-sectional area, is a binomial random variable with parameters N_A and g_A. Because g_A is typically small and we will aggregate over a large enough cross-sectional area to make the sum of the N_A's large, we approximate the random variable X_A by a Poisson random variable with parameter $N_A g_A$, which is both its mean and its variance.

We derive N_A by assuming that a gamma radiography machine can penetrate 16 cm of steel (SAIC claims that its various models of VACIS machines can penetrate between 9.5 and 16.5 cm [9]), which means that the aggregated machine measurements retain a signal-to-noise ratio of one if a standard $5 \times 10 \times 20$ cm lead brick is behind the steel. Consider two scenarios: in scenario 1, the lead brick is behind 16 cm of iron (the main element of steel), and hence $g_{A1} = e^{-5\mu_{lG} - 16\mu_{iG}}$ by (8), where $\mu_{lG}^{-1} = 1.544$ cm and $\mu_{iG}^{-1} = 2.424$ cm are the mean free paths of gamma rays in lead and iron, respectively [10]. The 200 aggregated measurements (the brick's cross-section is 200 cm^2) constitute a Poisson random variable X_{A1} with mean $200 N_A g_{A1}$, because the sum of Poisson

random variables is itself Poisson. In scenario 2, the lead is absent and $g_{A2} = e^{-16\mu_{iG}}$. The aggregated random variable is Poisson with mean $200N_A g_{A2}$. Now let $\Delta = X_{A2} - X_{A1}$ be the difference in the number of detected gamma rays between the two scenarios. Its mean is $200N_A(g_{A2} - g_{A1})$ and its standard deviation is $\sqrt{200N_A(g_{A1} + g_{A2})}$ because the two measurements are independent. Finally, we determine N_A so that the mean of Δ equals the standard deviation of Δ (this is what is meant by a signal-to-noise ratio of one), which yields

$$N_A = \frac{g_{A1} + g_{A2}}{200(g_{A2} - g_{A1})^2}. \tag{9}$$

A.4 False Positive Probabilities

We assume a fraction $f_T = 0.97$ of US-bound containers are trusted, meaning that they have not been flagged by the ATS. We view the testing system as having four stages: passive neutron monitoring (N), passive gamma monitoring (G), active radiography (A), and manual testing (M). For $i = \{N, G, A\}$, let f_i equal the false positive probability that a container generates an alarm for test i. Note that a weapon-free container can generate a false positive for three reasons: its contents, the natural background (in the case of passive testing), and measurement error.

The values of f_i are not set exogenously, but rather are determined by the test threshold parameters s_N, s_G and p_A, respectively, which are decision variables in our model. By the same reasoning as in (6), we have for $i = \{N, G\}$,

$$f_i = P\Big(X_i > s_i\Big), \tag{10}$$

where X_i is a normal random variable with mean μ_i and standard deviation $k_i\sqrt{\mu_i}$, where μ_i itself is the random variable $Z_i = \frac{A\epsilon_i C_i L}{4\pi r^2 v} + B_i$.

In contrast to passive testing, active testing can cause alarms for a variety of reasons that are independent of nuclear or radiological emissions. Consequently, we assume that $\tilde{f}_A = 0.05$ of actively tested containers set off an alarm regardless of the value of the threshold parameter

p_A, simply because an unexpected or mysterious object is seen [11]. In addition, an active testing alarm occurs if the container's contents are too dense for radiography to penetrate (e.g., shipments of metal objects, or certain agricultural shipments) or too difficult to decipher (e.g., the container contains a hodgepodge of items). Because containers filled with metal objects typically are 20 ft in length (to satisfy the weight limit), for modeling purposes we assume that a fraction $f_d = 0.1$ of containers are 20 ft in length and contain 24 cm of steel ($= 10\%$ packing fraction $\times$ 8 ft) along the direction measured by active testing. These containers are surrogates for all 20-ft and 40-ft dense or indecipherable containers, and the test threshold level p_A is applied only to these dense containers. By (8)-(9), each of these containers has a test measurement given by a Poisson random variable X_{An} with mean $N_A g_{An}$, where

$$g_{An} = e^{-\mu_{iG} r_n}, \tag{11}$$

$\mu_{iG}^{-1} = 2.424$ cm (iron) and $r_n = 24$ cm. We assume that an alarm occurs if the probability that the radiography measurement equals zero, which is $e^{-N_A g_{An}}$, is greater than the threshold p_A. Hence, the total false positive probability takes on one of two values:

$$f_A = \begin{cases} \tilde{f}_A & \text{if} \quad e^{-N_A g_{An}} \leq p_A; \\ \tilde{f}_A + f_d & \text{if} \quad e^{-N_A g_{An}} > p_A. \end{cases} \tag{12}$$

If $p_A = 1$ then containers never fail radiography because of their denseness, whereas a value of p_A near 0 provides a more aggressive strategy against dense containers (i.e., these containers fail active testing and are subsequently opened up because they were too dense to penetrate).

A.5 Detection Probability

There are three layers of protection to detect a weaponized container before it enters the port of embarkation: C-TPAT's certification system, mechanical container seals and the ATS software system. Recall that the weapon can be hidden in a 40-ft or 20-ft container from a certified or uncertified shipper. Define d_C to be the probability that a certified shipper would catch a terrorist

attempting to smuggle a nuclear weapon in one of its containers, d_S to be the probability that the mechanical seal sets off an alarm if a terrorist inserts the nuclear weapon into the container, and d_E to be the probability that the software system ATS would detect a container from an uncertified shipper that contained a nuclear weapon.

For $i = \{N, G, A\}$, let d_i be the probability that test i would detect a nuclear weapon. These probabilities depend on the testing decisions s_N, s_G and p_A, and the shielding thickness r_s. We assume that the lithium hydride shielding reduces the neutron and gamma emissions by the multiplicative factors f_{sN} and f_{sG}, respectively; e.g., if the neutron emissions are reduced by a factor of 100 then $f_{sN} = 0.01$. Since the neutron detector aggregates the neutrons detected over all energy levels, we are interested in the fraction of neutrons emitted by the weapon that is not absorbed by the shielding. We approximate this fraction by $f_{sN} = e^{-\mu_{sN} r_s}$ [1, 12], where μ_{sN}^{-1} is the mean free path of neutron absorption in lithium hydride. We set $\mu_{sN}^{-1} = 3.13$ cm, using the observation that $r_s = 20$ cm reduces the neutron emissions of the plutonium weapon in Fetter et al. by a factor of 600 [1]. By (8), we assume that the fraction of gamma rays undegraded in energy is $f_{sG} = e^{-\mu_{sG} r_s}$, where $\mu_{sG}^{-1} = 10.725$ cm is the mean free path of gamma rays in lithium hydride at 0.662 MeV for the plutonium weapon and μ_{sG}^{-1}=13.02 cm is the mean free path of gamma rays in lithium hydride at 1.001 MeV for the uranium weapon [10]. For $i = \{N, G\}$, we have from (5) that

$$d_i = P\Big(X_i > s_i\Big), \tag{13}$$

where X_i is a normal random variable with mean μ_i and standard deviation $k_i\sqrt{\mu_i}$, where $\mu_i = \frac{f_{si} \tan^{-1}\left(\frac{L}{2r}\right) A \epsilon_i S_i}{2\pi v r} + B_i$.

To derive d_A, we define X_{Aw} by (8)-(9), where for the plutonium weapon,

$$g_{Aw} = e^{-(1.5\mu_{pG} + 4\mu_{bG} + 6\mu_{tG} + 20\mu_{eG} + 2\mu_{aG} + 2\mu_{sG} r_s + \theta\mu_{iG}[244 - 42 - 2r_s])}, \tag{14}$$

where θ is the packing fraction; in equation (28), we change 1.5 cm to 2.46 cm, and 42 cm to 46 cm, for the uranium weapon (figure 2 of main text). To understand (14), note that the container,

which is 8 ft (or 244 cm) across, consists of the weapon (figure 2 of the main text), which is 42 cm in diameter and contains a 8.5 cm diameter empty core, a 1.5 cm thick layer (0.75 cm on each side of the weapon) of plutonium (with mean free path $\mu_{pG}^{-1} = 0.786$ cm; all mean free paths of gamma rays in (14) are at 1.3 MeV, which is the energy used in the newest VACIS machines [13], and are taken from [10]), 4 cm of beryllium (mean free path $\mu_{bG}^{-1} = 10.932$ cm), 6 cm of tungsten (mean free path μ_{tG}^{-1} =0.957 cm), 20 cm of high explosives (with mean free path $\mu_{eG}^{-1} = 9.185$ cm, calculated as in [14] from the composition of hydrogen, carbon, nitrogen and oxygen in the ratio 2:1:2:2 [1]), and 2 cm of aluminum (mean free path $\mu_{aG}^{-1} = 6.871$ cm), a layer of lithium hydride shielding of thickness r_s on both sides of the weapon ($\mu_{sG}^{-1} = 14.854$ cm), and a collection of identical iron objects ($\mu_{iG}^{-1} = 2.424$ cm) that have packing fraction θ. The packing fraction is given by

$$\theta = \frac{\left(2.95 \times 10^7 \text{ g} - \frac{4}{3}\pi[(r_s+21)^3 - 21^3] \text{ cm}^3 \text{ } 1.2 \text{ g/cm}^3\right)/7.8 \text{ g/cm}^3}{[7.25 \times 10^7 - \frac{4}{3}\pi(r_s+21)^3] \text{ cm}^3} \quad \text{for a } 40-\text{ft container,} \tag{15}$$

$$\theta = \frac{\left(2.95 \times 10^7 \text{ g} - \frac{4}{3}\pi[(r_s+21)^3 - 21^3] \text{ cm}^3 \text{ } 1.2 \text{ g/cm}^3\right)/7.8 \text{ g/cm}^3}{[3.62 \times 10^7 - \frac{4}{3}\pi(r_s+21)^3] \text{ cm}^3} \quad \text{for a } 20-\text{ft container,} \tag{16}$$

for the plutonium weapon; the 21 cm is changed to 23 cm in (15)-(16) for the uranium weapon (figure 2 of the main text). The numerator of (15) is the container weight limit (65k lbs) minus the weight of the lithium hydride shielding (assuming it is a concentric shell, where the density of the lithium hydride shielding is 1.2 g/cm^3 [1], which is about 50% higher than the density under normal conditions), divided by the density of steel; this ratio is the volume of steel that can be put into the container (ignoring the weight of the weapon, which is only 129 kg [1]). The denominator of (15) is the total container volume minus the volume of the shielded weapon. Equation (16) differs from (15) only in the calculation of the total container volume in the denominator.

There are two distinct portions of the container as it is scanned lengthwise: the weapon-

containing portion and the weapon-free portion. Because the outer radius of the tungsten layer is 10 cm [1] and the spatial resolution of gamma radiography is about 1 cm [9], we assume that radiography obtains $\pi 10^2 = 314$ weapon-containing measurements that are independent Poisson random variables with mean $N_A g_{Aw}$. The weapon-free portion of the container has $g_{An} = e^{-244\theta\mu_{iG}}$, since it is filled with identical iron objects according to the packing fraction θ. Because radiography generates nearly 3×10^5 weapon-free measurements from a 40-ft container, we can safely assume that g_{An} is correctly estimated during scanning. We allow the weapon to be detected via two approaches. First, as in (12), the container fails radiography if it is too dense, i.e., if $e^{-N_A g_{An}} > p_A$. Second, the weapon is detected if the 314 weapon-containing measurements differ significantly from the weapon-free measurements. For $j = \{n, w\}$, let $\tilde{X}_{Aj}$ be the sum of 314 independent samples from X_{Aj}, which itself is a Poisson random variable with mean $314 N_A g_{Aj}$. We assume that the weapon is detected if the aggregate measurement $\tilde{X}_{Aw}$ is less than the 0.05 tail of $\tilde{X}_{An}$; i.e., if the aggregate weapon measurement differs, at the 5% significance level, from the aggregate measurement we would expect to see from the weapon-free portion of the container. Taken together, we assume that

$$d_A = \begin{cases} 1 & \text{if} \quad e^{-N_A g_{An}} > p_A; \\ P(\tilde{X}_{Aw} < n^*) & \text{if} \quad e^{-N_A g_{An}} \leq p_A, \end{cases} \tag{17}$$

where

$$n^* = \min\left\{ n \;\middle|\; \sum_{k=0}^{n} \frac{(314 N_A g_{An})^k e^{-314 N_A g_{An}}}{k!} \geq 0.05 \right\}. \tag{18}$$

If the weapon is on a container of a certified shipper, then the detection probability (DP) for strategy YZ(a) is

$$\mathrm{DP}(\mathrm{YZ}(a)) = 1 - \prod_{i=\{C,S,N,G\}} (1 - d_i)(1 - a d_A) \quad \text{for Y} = \text{A, Z} = \text{E or D}, \tag{19}$$

$$\mathrm{DP}(\mathrm{YZ}(a)) = 1 - \prod_{i=\{C,S\}} (1 - d_i) \quad \text{for Y} = \text{U, Z} = \text{E or D}. \tag{20}$$

If the weapon is on a container of an uncertified shipper, then the detection probability is

$$\mathrm{DP}(\mathrm{YZ}(a)) = 1-(1-d_S)\Big[\prod_{i=\{E,N,G\}}(1-d_i)(1-ad_A)+d_E\prod_{i=\{N,G,A\}}(1-d_i)\Big] \text{ for Y = A, Z = E or D,} \tag{21}$$

$$\mathrm{DP}(\mathrm{YZ}(a)) = 1-(1-d_S)\Big[(1-d_E)+d_E\prod_{i=\{N,G\}}(1-d_i)(1-ad_A)\Big] \text{ for Y = U, Z = E or D.} \tag{22}$$

If inspection is carried out at both ports, then the analogs to (19)-(22), respectively, are

$$\mathrm{DP}(\mathrm{AB}(a)) = 1-\prod_{i=\{C,S\}}(1-d_i)\Big[\prod_{i=\{N,G\}}(1-d_i)(1-ad_A)\Big]^2 \text{ if certified,} \tag{23}$$

$$\mathrm{DP}(\mathrm{UB}(a)) = 1-\prod_{i=\{C,S\}}(1-d_i) \text{ if certified,} \tag{24}$$

$$\mathrm{DP}(\mathrm{AB}(a)) = 1-(1-d_S)\left[(1-d_E)\left(\prod_{i=\{N,G\}}(1-d_i)(1-ad_A)\right)^2+d_E\left(\prod_{i=\{N,G,A\}}(1-d_i)\right)^2\right] \text{ if uncertified,} \tag{25}$$

$$\mathrm{DP}(\mathrm{UB}(a)) = 1-(1-d_S)\left[(1-d_E)+d_E\left(\prod_{i=\{N,G\}}(1-d_i)(1-ad_A)\right)^2\right] \text{ if uncertified.} \tag{26}$$

B Congestion

The parametric-decomposition approach, which is used to estimate the probability distribution for the amount of time a container spends in the testing process at each port, is described in §B.1. The queueing networks at the port of embarkation and the port of debarkation are specified in §B.2 and §B.3, respectively. The values of the congestion parameters are given in Table 2. As explained in the main text, we ignore any congestion due to passive testing and we consider a single isolated ship that loads or unloads 3000 containers.

B.1 The Parametric-decomposition Approach

In the parametric-decomposition procedure [15], each queue in a network is characterized by five parameters: the arrival rate λ, the coefficient of variation (i.e., standard deviation divided by the

mean) of interarrival times c_a, the service rate μ for each server, the coefficient of variation of service times c_s, and the number of servers m.

In this subsection, we state the waiting time distribution at a generic queue with the five parameters given in the previous paragraph, as approximated by Whitt [15, 16]. In §B.2 and B.3, we derive four of the five parameters (the number of servers is a decision variable) for each of the queues in the testing networks, using subscripts to describe the port (E or D) and/or the test (A or M). The resulting waiting time distributions and service time distributions at each queue will then be combined to derive the total sojourn time distribution.

Let $\rho = \frac{\lambda}{m\mu}$ be the traffic intensity and $\beta = \sqrt{m}(1-\rho)$. We need to choose m so that $\rho < 1$ in order to maintain finite waiting times. Let W denote the steady-state waiting time in queue and D represent the conditional wait, given that the m servers are busy, i.e. $D = (W|W > 0)$. The expected waiting time in queue is approximated by

$$E(W) \approx \frac{c_a^2 + c_s^2}{2m\mu(1-\rho)(1+\sqrt{2\pi}\beta\Phi(\beta)e^{\beta^2/2})}, \tag{27}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal, and

$$P(W > 0) \approx \min\left\{1, \frac{1-\Phi\Big(2\beta/(1+c_a^2)\Big)}{\Big(1-\Phi(\beta)\Big)(1+\sqrt{2\pi}\beta\Phi(\beta)e^{\beta^2/2})}\right\}. \tag{28}$$

Using (27)-(28), we approximate the expected value of D by

$$E(D) \approx \frac{E(W)}{P(W>0)}, \tag{29}$$

and approximate the squared coefficient of variation of D by

$$c_D^2 \approx 2\rho - 1 + \frac{4(1-\rho)(2c_s^2+1)}{3(c_s^2+1)}. \tag{30}$$

The distribution of W is approximated by a point mass at zero with probability $1 - P(W > 0)$ using (28), and fitting a probability density function to D using (29)-(30), according to one of four cases, as described in equations (55)-(61) of [15].

The parametric-decomposition procedure approximates c_a at a downstream queue in terms of c_d, which is the coefficient of variation of the departure process of the upstream queue, and the fraction of containers exiting the upstream queue that go to the downstream queue. By [15], c_d^2 at a queue is approximated by

$$c_d^2 \approx 1 + (1-\rho^2)(c_a^2 - 1) + \frac{\rho^2}{\sqrt{m}}(c_s^2 - 1), \tag{31}$$

and if the departure process is randomly thinned (in our case, via testing) with probability p, then the resulting squared coefficient of variation of the interarrival times at the downstream queue is approximated by

$$pc_d^2 + 1 - p. \tag{32}$$

B.2 Port of Embarkation

Testing is performed at the gates leading into the terminal at the port of embarkation. We assume that US-bound containers arrive to the port of embarkation according to a Poisson process at rate λ_E. These containers arrive according to an appointment system with a one-hour time window, and typically arrive 6-8 hours before being loaded. If each truck driver behaves independently within the constraints of the appointment system, then the Poisson assumption is reasonably accurate [17]. To estimate λ_E, consider a US-bound ship that will load 3000 containers at the port of embarkation (see figure 1)). With four ship cranes each performing 30 moves/hr, this vessel will take 25 hours to load. Suppose containers are scheduled to arrive at constant rate λ_E from time 0 to time τ_2 (i.e., $\lambda_E = 3000/\tau_2$) and suppose loading starts at time $\tau_1 > 0$ and ends at time $\tau_1 + 25$. If we ignore variability, set the maximum waiting time for any container at 8 hours (i.e., $\tau_1 + 25 - \tau_2 = 8$), and assume that, to avoid ship-crane idleness, two hours worth of work (i.e., 240 containers) is at the ship crane before it starts loading (i.e., $3000\tau_1/\tau_2 = 240$), then $\tau_1 = 1.48$ hr, $\tau_2 = 18.48$ hr, and $\lambda_E = 162.3$/hr.

The arrival rate to active testing is

$$\lambda_{EA} = \lambda_E\Big[1 - f_T(1-a)\prod_{i=\{N,G\}}(1-f_i)\Big] \text{ for Strategy A,} \tag{33}$$

$$\lambda_{EA} = \lambda_E(1-f_T)\Big[1 - (1-a)\prod_{i=\{N,G\}}(1-f_i)\Big] \text{ for Strategy U.} \tag{34}$$

We assume $c_{aEA} = 1$, since a randomly thinned (i.e., containers randomly chosen with probability λ_{EA}/λ_E) Poisson process is also a Poisson process. We also assume $\mu_A = 20$/hr (i.e., three minutes to test a container [9]). Because the scan time is deterministic and the analysis of the scan is random, we assume that the service time is an Erlang (order two) random variable with $c_{sA}^2 = 0.5$. By (31), the squared coefficient of variation of the departure process from this queue is approximated by

$$c_{dEA}^2 \approx 1 + (1-\rho_{EA}^2)(c_{aEA}^2 - 1) + \frac{\rho_{EA}^2}{\sqrt{m_{EA}}}(c_{sA}^2 - 1), \tag{35}$$

where there are m_{EA} servers and the traffic intensity is $\rho_{EA} = \frac{\lambda_{EA}}{m_{EA}\mu_A}$.

Additional congestion will also be incurred by those containers that require manual inspection. Containers that pass both passive tests are manually tested only if they fail active testing, which occurs with probability f_A in (12). In contrast, containers that fail at least one of the two passive tests proceed from active testing to manual testing with probability $\tilde{f}_A + f_d$, regardless of the value of p_A in (12). That is, we assume that the passive testing failure is successfully diagnosed by active testing with probability $1 - \tilde{f}_A - f_d$, and the failure is diagnosed during manual testing otherwise. We denote the arrival rate of containers to manual inspection by λ_{EM}, which is given by

$$\lambda_{EM} = \lambda_E[1 - \prod_{i=\{N,G\}}(1-f_i)](\tilde{f}_A + f_d) + \lambda_E \prod_{i=\{N,G\}}(1-f_i)[f_T a + 1 - f_T]f_A \text{ for Strategy A,} \tag{36}$$

$$\lambda_{EM} = \lambda_E(1-f_T)\Big\{[1 - \prod_{i=\{N,G\}}(1-f_i)](\tilde{f}_A + f_d) + \prod_{i=\{N,G\}}(1-f_i)af_A\Big\} \text{ for Strategy U.} \tag{37}$$

By (32), the squared coefficient of variation of the interarrival times to manual testing is approximated by

$$c_{aEM}^2 = \frac{\lambda_{EM}}{\lambda_{EA}} c_{dEA}^2 + 1 - \frac{\lambda_{EM}}{\lambda_{EA}}, \tag{38}$$

where c_{dEA}^2 is given in (35).

Manual inspection is typically performed in teams. It takes five inspectors about three hours to completely empty and repack a container, and considerably less to open the container and peer inside. We assume that the mean service time is one hour ($\mu_A = 1$/hr) and that each of the m_{EM} servers represents a team of five inspectors. We assume that these service times are exponential (and hence $c_{sM}^2 = 1$), so that the probability that a manually-tested container is entirely emptied is about $e^{-3} = 0.05$.

For $i = \{A, M\}$, let the waiting time and the service time for queue i be denoted by W_{Ei} and S_{Ei}, respectively. Then the sojourn time T_E is given by

$$T_E = \begin{cases} 0 & \text{with probability } 1 - \frac{\lambda_{EA}}{\lambda_E}; \\ W_{EA} + S_{EA} & \text{with probability } \frac{\lambda_{EA} - \lambda_{EM}}{\lambda_E}; \\ W_{EA} + S_{EA} + W_{EM} + S_{EM} & \text{with probability } \frac{\lambda_{EM}}{\lambda_E}. \end{cases} \tag{39}$$

Because the waiting times and service times are independent, and the waiting times at various queues are assumed to be independent (e.g., [18]), the probability distribution of the steady-state total sojourn time T_E in the queueing network is approximated by a mixture (according to the probabilities in (39)) of the convolution of the waiting time and service time distributions at each queue.

B.3 Port of Debarkation

At the port of debarkation, passive monitoring is undertaken when the containers are on bombcarts just after the containers are taken off the ship, and active and manual testing is performed while the containers are on utility trucks, which pick up the containers after the bombcarts deposit them in the shipyard. We assume that three ship cranes, each working at rate 30/hr, are used to unload the

vessel, generating a container arrival rate of $\lambda_D = 90$/hr. The ship cranes work two shifts per day, and hence the 3000 containers from the ship are unloaded in about 33 hours, i.e. two days. The arrival rate λ_{DA} to active testing is again given by equations (33)-(34), but with the subscript D in place of E. If we view the ship unloading process as three servers working at 100% utilization (i.e., $\rho = 1$) with deterministic service times (i.e., $c_s = 0$), then equations (31)-(32) imply that the interarrival times of containers to active testing has squared coefficient of variation approximated by

$$c_{aDA}^2 \approx \frac{\lambda_{DA}}{\lambda_D}\Big[1 - \frac{1}{\sqrt{3}}\Big] + 1 - \frac{\lambda_{DA}}{\lambda_D}. \tag{40}$$

The service rates and coefficients of variation of service times for both active and manual testing are the same as at the port of embarkation. The arrival parameters λ_{DM} and c_{aDM} and the sojourn time T_D are given exactly as in (35)-(39), with the subscript D in place of E throughout.

C Costs

The total annual global cost $K(\text{YZ}(a))$ of strategy YZ(a) is given by

$$K(\text{YE}(a)) = k_{EP} + v_{EA}m_{EA} + v_{EM}m_{EM} \quad \text{for Y} = \text{A or U}, \tag{41}$$

$$K(\text{YD}(a)) = k_{DP} + k_{DA} + v_{DA}m_{DA} + v_{DM}m_{DM} \quad \text{for Y} = \text{A or U}, \tag{42}$$

$$K(\text{YB}(a)) = k_{DA} + \sum_{i=\{E,D\}} k_{iP} + v_{iA}m_{iA} + v_{iM}m_{iM} \quad \text{for Y} = \text{A or U}, \tag{43}$$

where, for i =E or D, k_{iP} is the annual cost of passive testing, k_{DA} is the annual cost of active testing at the port of debarkation that is independent of the number of active testers, and v_{ij} is the annual global cost per server for $j = A$ for active testing and $j = M$ for manual testing. We ignore the costs of C-TPAT, mechanical container seals and ATS, since these activities are assumed to already be in place. These seven cost parameters (see Table 3 for the values of the components of these parameters) take into account equipment and labor, and are assessed by the annual salary plus 0.2/yr times the cost of equipment.

For Strategy YZ(a), the passive testing cost is independent of Y and a, and only depends on Z (i.e., on where passive testing is performed). The cost of a passive portal monitor is about \$80k [11]. We assume that one portal detector is required per terminal. The 30 largest US seaports accounted for 99.5% of the imported containers in 2002 [19], and they have about 50 terminals in total. About 40 overseas ports ship to Pier 400 at Long Beach, and 105 ports ship to New York/New Jersey; the overlap between these ports is about 20, and some of these 125 distinct ports use several terminals. Hence, we estimate that there are about 150 terminals at ports of embarkation. We assume that each passive detector requires two full-time employees to operate each 8-hour shift per day, and each employee earns \$75k per year (we assume overseas monitors are manned by US employees, as dictated by the Container Security Initiative). Hence, our passive cost parameters are

$$k_{EP} = 150 \text{ terminals} \times \left[\frac{0.2}{\text{yr}}\Big(\frac{\$80\text{k}}{\text{terminal}}\Big) + \Big(6 \text{ workers} \times \frac{\$75\text{k}}{\text{worker} \cdot \text{yr}}\Big)\right] = \frac{\$69.9\text{M}}{\text{yr}}, \tag{44}$$

$$k_{DP} = 50 \text{ terminals} \times \left[\frac{0.2}{\text{yr}}\Big(\frac{\$80\text{k}}{\text{terminal}}\Big) + \Big(4 \text{ workers} \times \frac{\$75\text{k}}{\text{worker} \cdot \text{yr}}\Big)\right] = \frac{\$15.8\text{M}}{\text{yr}}. \tag{45}$$

The cost of an active portal monitor is about \$100k, although truck-mounted monitors are considerably more expensive [11]. We assume that an active tester requires three operators (each earning \$75k) per shift annually. Active testers scale with the number of ship cranes. Domestically, the Los Angeles/Long Beach ports have 135 ship cranes and handle 44% of imports, while Savannah has 13 cranes and handles 4.3% of the imports. Hence, we assume that the number of domestic ship cranes is approximately $\frac{135}{0.44} \approx \frac{13}{0.043} \approx 300$. We assumed three ship cranes were used at the port of debarkation in §B.3, and so we scale the number of active testers required at the port of debarkation up by a factor of 100 to estimate the domestic costs. The number of overseas-to-domestic ship cranes is assumed to be identical to the corresponding ratio for terminals, which is three. Hence, we scale the overseas active testers up by the factor of 300, which is three times

the domestic scale-up factor of 100. The annual cost per active tester at the two ports are

$$v_{EA} = 300 \times \left[\frac{0.2}{\text{yr}} \Big(\frac{\$100\text{k}}{\text{tester}} \Big) + \Big(\frac{9 \text{ workers}}{\text{tester}} \times \frac{\$75\text{k}}{\text{worker} \cdot \text{yr}} \Big) \right] = \frac{\$208.5\text{M}}{\text{yr}}, \tag{46}$$

$$v_{DA} = 100 \times \left[\frac{0.2}{\text{yr}} \Big(\frac{\$100\text{k}}{\text{tester}} \Big) + \Big(\frac{6 \text{ workers}}{\text{tester}} \times \frac{\$75\text{k}}{\text{worker} \cdot \text{yr}} \Big) \right] = \frac{\$47\text{M}}{\text{yr}}. \tag{47}$$

There are also active testing costs at the port of debarkation that are independent of the number of active testers. By the discussion of the logistics at the port of debarkation in §B.3, all actively-tested railbound containers require three crane movements by the tophandlers, compared to the traditional approach where railbound containers only required one movement by the tophandlers. If we let $p_r = 0.4$ denote the fraction of imported containers that are railbound, then the ratio of additional tophandlers required divided by the traditional (i.e., pre-inspection) number of tophandlers is

$$\frac{2p_r \frac{\lambda_{DA}}{\lambda_D}}{p_r + 2(1 - p_r)}. \tag{48}$$

Similarly, in the absence of inspection, only truckbound containers traveled on a utility truck, but now actively-tested railbound containers also require a trip in a utility truck. Therefore, the ratio of additional utility trucks required divided by the pre-inspection number of utility trucks is

$$\frac{p_r \frac{\lambda_{DA}}{\lambda_D}}{1 - p_r}. \tag{49}$$

Given the 3.4 tophandlers and 16.9 utility trucks per ship crane at Pier 400, we assume the three cranes used at the port of debarkation required 10 tophandlers and 51 utility trucks in the absence of inspection. Moreover, tophandlers cost about \$150k and utility trucks cost \$20k, and we assume the salary for the drivers are \$75k and \$50k per shift, respectively.

Tophandlers and utility trucks also scale with the number of ship cranes, with factors of 100 domestically and 300 overseas. Hence, the cost factor k_{DA} is given by

$$k_{DA} \;=\; 100 \times \left[10 \text{ tophandlers} \times \frac{2p_r \frac{\lambda_{DA}}{\lambda_D}}{2 - p_r} \left[\frac{0.2}{\text{yr}} \Big(\frac{\$150\text{k}}{\text{tophandler}} \Big) + \frac{2 \text{ workers}}{\text{tophandler}} \Big(\frac{\$75\text{k}}{\text{worker} \cdot \text{yr}} \Big) \right] \right.$$

$$
\begin{aligned}
&+ \quad 100 \times \left[51 \text{ utility trucks} \times \frac{p_r \frac{\lambda_{DA}}{\lambda_D}}{1 - p_r} \left[\frac{0.2}{\text{yr}} \left(\frac{\$20\text{k}}{\text{utility truck}} \right) + \frac{2 \text{ workers}}{\text{utility truck}} \left(\frac{\$50\text{k}}{\text{worker} \cdot \text{yr}} \right) \right] \right] \\
&= \quad \left(\frac{2p_r \frac{\lambda_{DA}}{\lambda_D}}{2 - p_r} \right) \frac{\$180\text{M}}{\text{yr}} + \left(\frac{p_r \frac{\lambda_{DA}}{\lambda_D}}{1 - p_r} \right) \frac{\$530.4\text{M}}{\text{yr}}.
\end{aligned}
$$

The parameter k_{DA} for Strategy YD(a) depends on Y and a via λ_{DA} in (33)-(34).

Each manual inspection team consists of five workers, each earning \$80k/yr. Since manual testers also scale with the number of cranes, we have

$$v_{EM} = 200 \times \left[\frac{3 \text{ teams}}{\text{tester}} \times \frac{\$400\text{k}}{\text{team} \cdot \text{yr}} \right] = \frac{\$360\text{M}}{\text{yr}}, \tag{50}$$

$$v_{DM} = 100 \times \left[\frac{2 \text{ teams}}{\text{tester}} \times \frac{\$400\text{k}}{\text{team} \cdot \text{yr}} \right] = \frac{\$80\text{M}}{\text{yr}}. \tag{51}$$

D Strategy Optimization and Evaluation

Each of the six classes of policies (AD,AE,AB,UD,UE,UB) contain six decision variables, which are listed in Table 4; we assume that these six variables take on the same value at each port in the two "B" policies. Hence, as noted in the main text, the US Government essentially has eight decisions: A or U; E, D, or B; and the six variables in Table 4. We evaluate each of these six classes of policies by choosing the decision variables to maximize the detection probability subject to a budget constraint of B dollars on the annual global cost, and congestion constraints at both ports. For the port of embarkation, we use the congestion constraint that at least 99% of containers make it through the testing process within 6 hours. This should lead to minimal impact on port efficiency and wreak little havoc on the loading plan. For the port of debarkation, we assume that at least 95% of containers make it through the testing process within 4 hours. Here, we are holding up individual containers and not an entire ship, and so we can use a lower service level.

The terrorists have three decisions: the shielding thickness r_s, a 20-ft or 40-ft container, and a certified or uncertified shipper. The optimization problem corresponding to the Stackelberg game [20], where the US Government makes its decisions before the terrorists, is

$$\max_{\{a,m_A,m_M,s_N,s_G,p_A\}} \quad \min_{\substack{r_s \\ \text{20-ft or 40-ft} \\ \text{certified or uncertified}}} \quad \text{DP}(\text{YZ}(a)) \tag{52}$$

$$\text{subject to} \quad K(\text{YZ}(a)) \leq B, \tag{53}$$

$$P(T_E > \ 6 \text{ hr}) \leq 0.01, \tag{54}$$

$$P(T_D > \ 4 \text{ hr}) \leq 0.05, \tag{55}$$

$$\frac{4}{3}\pi\bigg(\Big(r_s + 21\Big)^3 - 21^3\bigg)1.2 \leq 5.9 \times 10^6, \tag{56}$$

where DP(YZ(a)) is defined in (19)-(26), K(YZ(a)) is defined in (41)-(43), and T_E and T_D are defined in (39), where it is understood that, with probability one, $T_E = 0$ if $Z = D$ and $T_D = 0$ if $Z = E$. Constraint (56) (expressed in grams) is based on the assumption that the ATS would not detect the shielding as long as the weight of the lithium hydride shielding is no more than 13k lbs, which is 20% of the container weight limit (we omit the mass of the weapon, which is only 129 kg [1]). The left side of (56) assumes the shield is a concentric shell, and follows from (14); in (56), we change 21 cm to 23 cm for the uranium weapon (Figure 2 of the main text). Constraint (56) reduces to $r_s \leq 84.8$ cm for the plutonium weapon; the corresponding outer radius of shielding is $84.8 + 21$ =105.8 cm, which is slightly less than the space constraint of 122 cm. Similarly, equation (56) reduces to $r_s \leq 82.9$ cm for the uranium weapon. By re-optimizing for different values of B in (53), we generate the cost vs. detection probability curves in the main text.

The initial input to the optimizer is a set of values for the decision variables a, s_N, s_G, p_A, from which we compute the minimum cost combination of m_A and m_M that will satisfy the maximum waiting time constraints (54)-(55). The optimizer then computes the numerical gradient of the objective function with the initial set of decision variable values. It uses the gradient to determine how it should alter the decision variables to produce the greatest increase to the objective function, attempting progressively smaller steps of the decision variables in that direction until it finds a new set of values that still satisfies all the constraints and increases the objective function.

The optimizer stops once the gradient is too small (indicating that the function is barely changing) or when it can no longer find a new set of values that both satisfies the constraints and increases the objective function.

E Alternative Modes of Testing

In this section, we discuss alternative uses of existing technologies as well as technologies under development. The more promising options are considered in §8 of the main text.

E.1 Passive Testing at the Ports

Portal monitors are only one option for passive neutron and gamma testing at the ports of embarkation and debarkation. Handheld devices have considerably higher false positive probabilities and are much slower than portal monitors [2], making them impractical for front-line testing. However, they can be quite helpful in the investigation of containers that fail either passive or active testing.

Passive monitors can also be put on the ship cranes. One possibility is to put a sensor on the spreader bar that hovers over the container as it is handled. In this case, t is about 30 sec, r is about 1.5 m, and A_s and ϵ_s are similar to handheld devices. However, the results of an experiment in Norfolk, VA were not promising, due partly to the fluctuations in background radiation as the sensor moved in and out of the ship to get the container. Even if this serious technical issue can be overcome, another key hurdle is the durability of these monitors in such a harsh environment. Suppose the time between failures for a monitor on the beam is exponentially distributed with parameter τ (i.e, a mean of τ^{-1} hr). Because it is too expensive to stop the ship crane to fix the monitor during the crane's two shifts of operations at a port of debarkation, the monitor would be repaired during the third shift. Under this assumption, the fraction of time over the two shifts that the monitor is operating is given by $\frac{1-e^{-16\tau}}{16\tau}$. Hence, the detection probability and the false positive

probability are both decreased by this factor.

A more promising option is to put the detector on the scaffold of the ship crane. In this case, t will be shorter (perhaps 5 sec), and A, ϵ_N and ϵ_G would be similar to the values in §A.1-§A.2 for a portal monitor. The monitor's reliability would be higher (i.e., smaller τ, because the scaffold undergoes less stress than the spreader bar) and its false positive rate would be lower (less background fluctuations). However, given that passive portal monitoring does not significantly impact port congestion, there does not appear to be a compelling need to pursue these difficult options.

E.2 Elongated Portal Passive Testing

Rather than driving through the portal monitor at velocity v, we assume that a container takes 10 sec to drive into the portal and then stays there for t sec, generating emissions at the detector of $\frac{f_{si} A \epsilon_i S_i t}{4\pi r^2}$. We create a 3-stage queueing network by adding a queue in front of the two-stage network in §B.3. The Poisson arrival rate to this queue is $\lambda_E = 162.3$/hr, the service times are deterministic with duration $\mu_p^{-1} = 10 + t$ sec, and there are m_{EP} servers. There are two additional decision variables, the testing time (t) and the number of servers (m_{EP}). Because the testing process is highly automated, we assume that two workers can run this operation, regardless of the number of machines.

E.3 Passive Testing inside Containers

Passive monitors can also be put inside containers. Given the low false positive probability of passive neutron testing, these sensors might be able to exploit the week-long trip to improve the signal-to-noise ratio of passive neutron testing at the port of debarkation. The steel container walls, which are $r_c = 0.2$ cm thick, reduce the background neutron source at the detector by the factor $e^{-\mu_{iN} r_c}$, where $\mu_{iN}^{-1} = x$ cm is the mean free path of neutrons in steel. Hence, the true cumulative

background emissions is a normal random variable B_N with mean $A\epsilon_N b_N t$ and standard deviation $\sqrt{A\epsilon_N b_N t}$, where $A = 0.02$ and $\epsilon_N = 0.05$ (similar to handheld devices [1]) and $t = 7$ days. The true weapon emissions at the detector would be $\frac{A\epsilon_N e^{-\mu_{iN} r_c} S_N t}{4\pi r^2}$, where $r = 1.2$ m (the halfwidth of a container). The remainder of the model remains unchanged, including the assumption that the typical weaponless container has no neutron emissions (i.e., $C_N = 0$ with probability one). Note that containers far from direct sunlight may receive a smaller background radiation rate than b_N, which causes us to underestimate the efficacy of this option. The cost of each sensor is assumed to be \$50, and the 0.2/yr equipment factor is applied.

E.4 X-ray Radiography

X-ray radiography is capable of penetrating up to 41 cm of steel [21]. However, this increased penetration is achieved with a 9-MeV x-ray source, which requires considerable safety precautions. Consequently, it is practical to only use these \$1.2M machines for containers that cannot be penetrated by gamma radiography. We consider a strategy where all dense containers that are actively tested are routed to x-ray radiography, and all other actively tested containers are routed to gamma radiography. We further assume that the fraction $\tilde{f}_A$ of containers that cause false positives are all processed by gamma radiography. If we assume that the service time characteristics of gamma radiography are the same as x-ray radiography and denote the service time and waiting time at x-ray radiography at the embarkation port by S_{EX} and W_{EX}, respectively, then equation (19) is replaced by

$$T_E = \begin{cases} 0 & \text{with probability } 1 - \frac{\lambda_{EA}}{\lambda_E}; \\ W_{EA} + S_{EA} & \text{with probability } \frac{(1-f_d-\tilde{f}_A)\lambda_{EA}}{\lambda_E}; \\ W_{EX} + S_{EX} & \text{with probability } \frac{f_d \lambda_{EA}}{\lambda_E}; \\ W_{EA} + S_{EA} + W_{EM} + S_{EM} & \text{with probability } \frac{\tilde{f}_A \lambda_{EA}}{\lambda_E}. \end{cases} \tag{57}$$

To estimate how the x-ray equipment would perform, we replace 16 cm by 41 cm in g_{A1} and g_{A2} in equation (9) and calculate a new value of N_A. In this and all other calculations to find the false

positive probability and detection probability for active testing, we use the following mean free paths at 9MeV: lithium hydride (45.0 cm), beryllium (31.6 cm), explosives (24.4 cm), aluminum (15.6 cm), iron (4.26 cm), lead (1.83 cm), tungsten (1.13 cm), plutonium (0.974 cm).

E.5 Networked Active Testing

Active testing throughput can be increased by electronically transmitting the scanned images and allowing multiple scans to be analyzed simultaneously, thereby decoupling scan production and scan analysis. To analyze the congestion resulting from this option, we consider a 3-stage queueing network consisting of scan production ($\mathrm{A_p}$), scan analysis ($\mathrm{A_a}$), and manual testing (M). The arrival parameters λ_{EA_p}, $c^2_{aEA_p}$, λ_{DA_p} and $c^2_{aDA_p}$ are identical to λ_{EA}, c^2_{aEA}, λ_{DA} and c^2_{aDA} respectively, in §B.2 and §B.3. The service rate at the scan production queue is $\mu_{A_p} = 60$/hr and the service times are assumed to be deterministic ($c_{sA_p} = 0$). Since all scans are analyzed, $\lambda_{iA_a} = \lambda_{iA_p}$ for $i =$E or D, and equations (31)-(32) imply that

$$c^2_{aiA_a} \approx 1 + (1 - \rho^2_{iA_p})(c^2_{aiA_p} - 1) - \frac{\rho^2_{iA_p}}{\sqrt{m_{iA_p}}} \quad \text{for} \quad \mathrm{i = E \text{ or } D}. \tag{58}$$

We assume that $\mu_A = 30$/hr, so that the mean total active testing time is the same as in the base case (i.e., $\mu_A^{-1} = \mu_{A_p}^{-1} + \mu_{A_a}^{-1}$). Similarly, we assume $c^2_{sA_a} = \frac{9}{8}$ (for simplicity, S_{iA_a} in (60) is assumed to be exponential), so that the squared coefficient of variation of the total testing time is the same as in the base case (i.e., $c^2_{sA_a} = \frac{\mu^2_{A_a}}{\mu^2_A} c^2_{sA}$). The arrival rates to manual inspection, λ_{EM} and λ_{DM}, are given by (36)-(37), and

$$c^2_{aiM} \approx \frac{\lambda_{iM}}{\lambda_{iA_p}} c^2_{diA_a} + 1 - \frac{\lambda_{iM}}{\lambda_{iA_p}} \quad \text{for} \quad \mathrm{i = E \text{ or } D}. \tag{59}$$

The service parameters μ_M and c_{sM} of manual testing are the same as in the base case. The total sojourn time is given by, for $i =$E or D,

$$T_i = \begin{cases} 0 & \text{with probability } 1 - \frac{\lambda_{iA_p}}{\lambda_i}; \\ W_{iA_p} + S_{iA_p} + W_{iA_a} + S_{iA_a} & \text{with probability } \frac{\lambda_{iA_p} - \lambda_{iM}}{\lambda_i}; \\ W_{iA_p} + S_{iA_p} + W_{iA_a} + S_{iA_a} + W_{iM} + S_{iM} & \text{with probability } \frac{\lambda_{iM}}{\lambda_i}. \end{cases} \tag{60}$$

We assume that each of the servers at scan production consist of an active tester ($100k each) and two operators ($75k/yr each), and each server at scan analysis consists of one inspector ($75k/yr each).

F Radiological

There are seven important radioisotopes [22] that can be put into radiological dispersal devices (RDDs), or so-called dirty bombs: ^{241}Am (found in smoke detectors), ^{252}Cf, ^{137}Cs (spent nuclear fuel and brachytherapy), ^{60}Co (nuclear reactors, food radiation), ^{192}Ir (cameras), ^{238}Pu (thermal generators), and ^{90}Sr (nuclear waste, television tubes). There are three main kinds of ionizing radiation: alpha particles, which only travel 4 cm but are harmful if inhaled; beta particles, which travel 2 m and can cause skin burns; and gamma rays and x-rays, which pose serious risks. ^{241}Am, ^{252}Cf and ^{238}Pu only emit alpha particles, and so can be easily handled by terrorists, whereas the other four isotopes can be fatal if handled in sufficient quantities.

Levi [22, 23] estimates that 2 Ci of ^{137}Cs (an amount found in many medical gauges) in 10 pounds of TNT would cause a km^2 (40 city blocks) to exceed EPA radiological standards, whereas a Russian study [24] states that between 50 and 10k Ci of ^{137}Cs in 50 kg of TNT could cause the population in a 1 km^2 area to seek shelter. The discrepancy between these studies could be due to the different safety thresholds used, the fact that the population in the Russian analysis takes protective measures, different amounts of explosives, and different atmospheric models (the Russian model is more complex). For concreteness, we consider a radiological source of 10 Ci (0.11 g) of ^{137}Cs.

Unlike fissile material, the emissions from dirty bomb materials are not due to the fission process. Consequently, radioactive material will be detected by its emitted gamma rays. Ten Ci of ^{137}Cs, if unshielded, generates 3.6×10^{12} gamma rays/sec, of which 46.92% is at 0.662 meV. Hence, the peak gamma-ray emission rate is 1.69×10^{12} rays/sec, compared with 600 or 30 gamma

rays/sec for the nuclear sources considered in the main text. To reduce the radiological gamma-ray emissions to the mean plus 3 standard deviations of the background, $A\epsilon_G b_G L/v + 3\sqrt{A\epsilon_G b_G L/v}$, requires 8.23 cm of tungsten shielding. The corresponding g_{Aw} term in (14) is $e^{-16.46\mu_{tG}} = 3.4 \times 10^{-8}$, which is about 3 logs smaller than the g_{An} term in (17)-(18) for a 20-ft container, and about 5 logs smaller for a 40-ft container. Consequently, 10 Ci of ^{137}Cs is at least as easy to detect as a nuclear weapon: if enough shielding is used to hide the gamma-ray emissions, then gamma radiography will detect the radiological source if the surrounding steel items can be penetrated.

G Supplementary Computational Results

Tables 5-8 provides the detailed solutions for the points on the base-case curves in Figs. 4a-4d, respectively, of the main text.

References

[1] S. Fetter, V. A. Frolov, M. Miller, R. Mozley, O. F. Prilutsky, S. N. Rodionov, R. Z. Sagdeev, *Science & Global Security* **1**, 225 (1990).

[2] P. Beck, Paper OEFZS-G-0005, Austrian Research Centers, Seibersdorf (2000).

[3] E. F. Plechaty, J. R. Kimlinger, UCRL-50400, Vol. 14, Lawrence Livermore National Laboratory (July 4, 1976).

[4] http://www.tsasystems.com/products/portals/vm-250.php.

[5] S. Fetter, T. B. Cochran, L. Grodzins, H. L. Lynch, M. S. Zucker, *Science* **248**, 828 (1990).

[6] http://www.csupomona.edu/˜pbsiegel/www/nuclear.html.

[7] http://www.eas.asu.edu/holbert/eee460/HealthPhysics.pdf.

[8] http://physics.nist.gov/PhysRefData/XrayMassCoef/cover.html.

[9] R. D. Richardson, V. V. Verbinski, V. J. Orphan, *Port Technology Int'l.*, 83 (March 4, 2002).

[10] http://physics.nist.gov/PhysRefData/Xcom/Text/XCOM.html.

[11] S. Schiesel, *N.Y. Times*, Section E, p. 1 (March 20, 2003).

[12] R. L. Maughan, M. Yudelev, J. D. Forman, S. B. Williams, D. Gries, T. M. Fletcher, W. Chapman, E. J. Blosser, T. Horste, *Med. Phys.* **28**, 1006 (2001).

[13] M. P. Snell, *Port Technology Int'l.*, 83 (date).

[14] http://www.tpub.com/doenuclearphys/nuclearphysics40.htm.

[15] W. Whitt, *Bell System Tech. J.* **62**, 2779 (1983).

[16] W. Whitt, *Production Operations Mgmt.* **2**, 114 (1993).

[17] E. Cinlar, in P. A. W. Lewis, Ed., *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, Wiley, NY (1972).

[18] J. M. Harrison, V. Nguyen, *Queueing Systems* **6**, 1 (1990).

[19] http://www.aapa-ports.org/industryinfo/statistics.htm.

[20] R. Gibbons, *Game Theory for Applied Economists*, Princeton University Press, Princeton, NJ (1992).

[21] http://www.heimanncargovision.com/shockedpage.html.

[22] C. D. Ferguson, T. Kazi, J. Perera, Paper No. 11, Center for Nonproliferation Studies, Monterey Institute of International Studies (2003).

[23] M. A. Levi, H. C. Kelly, *Scientific American* 76 (Nov. 2002).

[24] L. Bolshov, R. Arutyunyan, O. Pavlovsky, *High-Impact Terrorism: Proceedings of a Russian-American Workshop*, pg 137, National Academy of Sciences (2002).

Parameter	Description	Value	Reference
L	Container length	20 or 40 ft	See text
v	Velocity during passive testing	2.2 m/sec	See text
A	Area of radiation detector	0.3 m^2	[1]
r	Distance of radiation detector	2 m	[2]
S_N	Neutron source	400k neutrons/sec	[1]
ϵ_N	Efficiency of neutron detector	0.14	[1]
b_N	Mean neutron background rate	50 neutrons/m^2·sec	[1]
e^{c_N}	Median container neutron emissions	0	[2]
$e^{\sigma_{cN}}$	Dispersion of container neutron emissions	1	[2]
k_N	Neutron noise factor	2.81	(2)-(3)
S_G	Gamma source	600 gamma rays/sec	[1]
ϵ_G	Efficiency of gamma detector	0.70	[1]
b_G	Mean gamma background rate	1400 gamma rays/m^2·sec	[1]
e^{c_G}	Median container gamma emissions	2.63 gamma rays/sec	(6)-(7)
$e^{\sigma_{cG}}$	Dispersion of container gamma emissions	43.63	(6)-(7)
k_G	Gamma noise factor	0.146	(4)-(5)
N_A	Effective radiography emissions	4.14 gamma rays/cm^2	(8)-(9)
f_T	Fraction of trusted containers	0.97	See text
$\tilde{f}_A$	Fraction of containers alarming active test	0.05	[11]
f_d	Fraction of dense containers	0.1	(12)
r_n	Thickness of metal in dense containers	24 cm	See text
d_C	Detection probability of certification	0.2	See text
d_S	Detection probability of seals	0.05	See text
d_E	Detection probability of ATS	0.05	See text

Table 1: Values for detection-modeling parameters.

Parameter	Description	Value	Reference
λ_E	Embarkation truck arrival rate	162/hr	See text
c_{aEA}	cv of interarrival times at embarkation, active test	1	See text
μ_A	Active testing rate	20/hr	[9]
c_{sA}	Coefficient of variation (cv) of active test times	$\sqrt{0.5}$	See text
μ_M	Manual testing rate	1/hr	See text
c_{sM}	Coefficient of variation (cv) of manual test times	1	See text
λ_D	Debarkation truck arrival rate	90/hr	See text

Table 2: Values for congestion parameters. All other congestion parameter values are derived from other parameters and decision variables.

Description	Value	Reference
Fraction of railbound containers (p_r)	0.4	See text
Number of US terminals	50	[19]
Number of overseas terminals	150	See text
Cost of passive tester	\$80k	[11]
Number of employees per passive tester	2	See text
Employee salary at active testing	\$75k/yr	See text
Cost of gamma radiography machine	\$100k	[11]
Number of employees per active tester	3	See text
Employee salary at active testing	\$75k/yr	See text
Cost of tophandler	\$150k	See text
Salary of tophandler operator	\$75k/yr	See text
Cost of utility truck	\$20k	See text
Salary of utility-truck driver	\$50k/yr	See text
Salary of manual tester	\$75k/yr	See text
Cost of passive monitor on a seal	\$50	See text
Cost of electronic tamper-resistant seal	\$100	See text
Number of containers worldwide	12M	See text
Cost of x-ray radiography machine	\$1.2M	See text

Table 3: Cost parameters.

Parameter	Description
a	Fraction active testing
s_N	Neutron threshold level
s_G	Gamma threshold level
p_A	Radiography threshold probability
m_A	Number of active testers
m_M	Number of manual testing teams

Table 4: Decision variables.

L	Cert.	Strat.	DP	Budget	Cost	d_N	d_A	a	m_{EA}	m_{EM}	m_{DA}	m_{DM}
20	U	EA	0.108	650	638	0.00	1.0	0.01	1	1	0	0
20	U	EA	0.108	800	638	0.00	1.0	0.01	1	1	0	0
20	U	EA	0.145	1000	998	0.00	1.0	0.05	1	2	0	0
20	U	EA	0.145	1200	998	0.00	1.0	0.05	1	2	0	0
20	U	EA	0.179	1500	1358	0.00	1.0	0.09	1	3	0	0
20	U	EA	0.216	2000	1927	0.00	1.0	0.13	2	4	0	0
20	U	EA	0.289	3000	2855	0.00	1.0	0.21	3	6	0	0
20	U	EA	0.399	4000	3935	0.00	1.0	0.33	3	9	0	0
20	U	EA	0.475	5000	4864	0.00	1.0	0.42	4	11	0	0
20	U	DA	0.106	160	160	0.00	1.0	0.01	0	0	1	1
20	U	DA	0.138	240	176	0.00	1.0	0.05	0	0	1	1
20	U	DA	0.207	320	289	0.00	1.0	0.12	0	0	1	2
20	U	DA	0.276	480	448	0.00	1.0	0.20	0	0	2	3
20	U	DA	0.345	640	561	0.00	1.0	0.27	0	0	2	4
20	U	DA	0.474	800	783	0.00	1.0	0.42	0	0	2	6
20	U	DA	0.542	1000	989	0.00	1.0	0.49	0	0	4	7
20	U	DA	0.807	1500	1482	0.00	1.0	0.79	0	0	5	11
20	U	DA	1.000	2000	1814	0.00	1.0	1.00	0	0	5	14
20	U	BA	0.118	800	799	0.00	1.0	0.01	1	1	1	1
20	U	BA	0.118	1000	799	0.00	1.0	0.01	1	1	1	1
20	U	BA	0.178	1250	1174	0.00	1.0	0.05	1	2	1	1
20	U	BA	0.189	1500	1257	0.00	1.0	0.05	1	2	1	2
20	U	BA	0.253	1750	1633	0.00	1.0	0.09	1	3	1	2
20	U	BA	0.259	2000	1995	0.00	1.0	0.09	1	4	1	2
20	U	BA	0.380	3000	2885	0.00	1.0	0.17	3	5	1	3
20	U	BA	0.496	4000	3767	0.00	1.0	0.25	3	7	2	4
20	U	BA	0.615	5000	4968	0.00	1.0	0.35	3	10	2	5

Table 5: Solutions corresponding to points in Fig. 4a of main text. Under Cert. column, C is uncertified and U is certified. We report d_N and d_A in lieu of s_N and p_A. For all scenarios, $d_G = 0.00$ and $r_s = 84.8$ cm. Cost and Budget figures are in millions of dollars.

L	Cert.	Strat.	DP	Budget	Cost	d_N	d_A	a	m_{EA}	m_{EM}	m_{DA}	m_{DM}
20	C	EA	0.264	650	638	0.03	1.0	0.00	1	1	0	0
20	C	EA	0.264	800	638	0.03	1.0	0.00	1	1	0	0
20	C	EA	0.280	1000	998	0.00	1.0	0.05	1	2	0	0
20	C	EA	0.280	1200	998	0.00	1.0	0.05	1	2	0	0
20	C	EA	0.309	1500	1358	0.00	1.0	0.09	1	3	0	0
20	C	EA	0.340	2000	1927	0.00	1.0	0.13	2	4	0	0
20	C	EA	0.401	3000	2855	0.00	1.0	0.21	3	6	0	0
20	C	EA	0.494	4000	3935	0.00	1.0	0.33	3	9	0	0
20	C	EA	0.558	5000	4864	0.00	1.0	0.42	4	11	0	0
20	C	DA	0.247	160	160	0.00	1.0	0.01	0	0	1	1
20	C	DA	0.275	240	176	0.00	1.0	0.05	0	0	1	1
20	C	DA	0.333	320	289	0.00	1.0	0.12	0	0	1	2
20	C	DA	0.391	480	448	0.00	1.0	0.20	0	0	2	3
20	C	DA	0.449	640	561	0.00	1.0	0.27	0	0	2	4
20	C	DA	0.557	800	783	0.00	1.0	0.42	0	0	2	6
20	C	DA	0.614	1000	989	0.00	1.0	0.49	0	0	4	7
20	C	DA	0.837	1500	1482	0.00	1.0	0.79	0	0	5	11
20	C	DA	1.000	2000	1814	0.00	1.0	1.00	0	0	5	14
20	C	BA	0.259	800	800	0.01	1.0	0.00	1	1	1	1
20	C	BA	0.287	1000	808	0.03	1.0	0.00	1	1	1	1
20	C	BA	0.335	1250	1182	0.06	1.0	0.00	1	2	1	1
20	C	BA	0.344	1500	1474	0.07	1.0	0.00	2	2	1	2
20	C	BA	0.376	1750	1635	0.09	1.0	0.00	1	3	1	2
20	C	BA	0.397	2000	1850	0.11	1.0	0.00	2	3	1	2
20	C	BA	0.499	3000	2892	0.19	1.0	0.00	3	5	1	3
20	C	BA	0.593	4000	3774	0.27	1.0	0.00	3	7	2	4
20	C	BA	0.677	5000	4817	0.35	1.0	0.00	4	9	2	5

Table 6: Solutions corresponding to points in Fig. 4b of main text. Under Cert. column, C is uncertified and U is certified. We report d_N and d_A in lieu of s_N and p_A. For all scenarios, $d_G = 0.00$ and $r_s = 84.8$ cm. Cost and Budget figures are in millions of dollars.

L	Cert.	Strat.	DP	Budget	Cost	d_N	d_A	a	m_{EA}	m_{EM}	m_{DA}	m_{DM}
40	U	EA	0.182	650	638	0.00	1.0	0.09	1	1	0	0
40	U	EA	0.182	800	638	0.00	1.0	0.09	1	1	0	0
40	U	EA	0.184	1000	846	0.00	1.0	0.10	2	1	0	0
40	U	EA	0.184	1200	846	0.00	1.0	0.10	2	1	0	0
40	U	EA	0.294	1500	1415	0.00	1.0	0.22	3	2	0	0
40	U	EA	0.296	1600	1566	0.00	1.0	0.22	2	3	0	0
40	U	EA	0.400	2000	1983	0.00	1.0	0.34	4	3	0	0
40	U	EA	0.616	3000	2912	0.00	1.0	0.57	5	5	0	0
40	U	EA	0.754	4000	3840	0.00	1.0	0.73	6	7	0	0
40	U	EA	0.983	5000	4977	0.00	1.0	0.98	8	9	0	0
40	U	DA	0.106	160	159	0.00	1.0	0.01	0	0	1	1
40	U	DA	0.273	240	239	0.00	1.0	0.19	0	0	1	1
40	U	DA	0.276	320	288	0.00	1.0	0.20	0	0	2	1
40	U	DA	0.479	480	464	0.00	1.0	0.42	0	0	2	2
40	U	DA	0.580	640	639	0.00	1.0	0.53	0	0	3	3
40	U	DA	0.690	800	739	0.00	1.0	0.66	0	0	4	3
40	U	DA	0.897	1000	965	0.00	1.0	0.89	0	0	5	4
40	U	DA	1.000	1500	1094	0.00	1.0	1.00	0	0	5	5
40	U	BA	0.120	800	799	0.00	1.0	0.01	1	1	1	1
40	U	BA	0.263	1200	1044	0.00	1.0	0.10	2	1	1	1
40	U	BA	0.447	1600	1583	0.00	1.0	0.22	2	2	2	2
40	U	BA	0.452	2000	1944	0.00	1.0	0.22	2	3	2	2
40	U	BA	0.731	3000	2949	0.00	1.0	0.45	4	4	3	3
40	U	BA	0.856	4000	3940	0.00	1.0	0.60	5	6	3	3
40	U	BA	0.971	5000	4939	0.00	1.0	0.82	7	7	4	4

Table 7: Solutions corresponding to points in Fig. 4c of main text. Under Cert. column, C is uncertified and U is certified. We report d_N and d_A in lieu of s_N and p_A. For all scenarios, $d_G = 0.00$ and $r_s = 84.8$ cm. Cost and Budget figures are in millions of dollars.

L	Cert.	Strat.	DP	Budget	Cost	d_N	d_A	a	m_{EA}	m_{EM}	m_{DA}	m_{DM}
40	C	EA	0.311	650	638	0.00	1.0	0.09	1	1	0	0
40	C	EA	0.311	800	638	0.00	1.0	0.09	1	1	0	0
40	C	EA	0.313	1000	846	0.00	1.0	0.10	2	1	0	0
40	C	EA	0.313	1200	846	0.00	1.0	0.10	2	1	0	0
40	C	EA	0.406	1500	1415	0.00	1.0	0.22	3	2	0	0
40	C	EA	0.408	1600	1566	0.00	1.0	0.22	2	3	0	0
40	C	EA	0.495	2000	1983	0.00	1.0	0.34	4	3	0	0
40	C	EA	0.677	3000	2912	0.00	1.0	0.57	5	5	0	0
40	C	EA	0.793	4000	3840	0.00	1.0	0.73	6	7	0	0
40	C	EA	0.986	5000	4977	0.00	1.0	0.98	8	9	0	0
40	C	DA	0.247	160	159	0.00	1.0	0.01	0	0	1	1
40	C	DA	0.388	240	239	0.00	1.0	0.19	0	0	1	1
40	C	DA	0.391	320	288	0.00	1.0	0.20	0	0	2	1
40	C	DA	0.561	480	464	0.00	1.0	0.42	0	0	2	2
40	C	DA	0.646	640	639	0.00	1.0	0.53	0	0	3	3
40	C	DA	0.739	800	739	0.00	1.0	0.66	0	0	4	3
40	C	DA	0.913	1000	965	0.00	1.0	0.89	0	0	5	4
40	C	DA	1.000	1500	1094	0.00	1.0	1.00	0	0	5	5
40	C	BA	0.259	800	799	0.00	1.0	0.01	1	1	1	1
40	C	BA	0.379	1200	1044	0.00	1.0	0.10	2	1	1	1
40	C	BA	0.534	1600	1583	0.00	1.0	0.22	2	2	2	2
40	C	BA	0.538	2000	1944	0.00	1.0	0.22	2	3	2	2
40	C	BA	0.773	3000	2949	0.00	1.0	0.45	4	4	3	3
40	C	BA	0.879	4000	3940	0.00	1.0	0.60	5	6	3	3
40	C	BA	0.975	5000	4939	0.00	1.0	0.82	7	7	4	4

Table 8: Solutions corresponding to points in Fig. 4d of main text. Under Cert. column, C is uncertified and U is certified. We report d_N and d_A in lieu of s_N and p_A. For all scenarios, $d_G = 0.00$ and $r_s = 84.8$ cm. Cost and Budget figures are in millions of dollars.

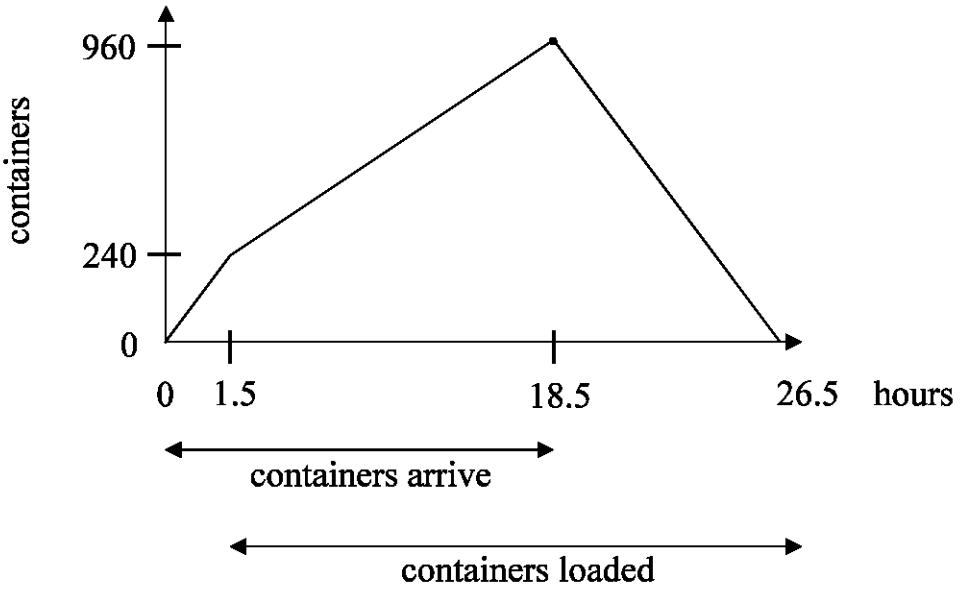

Figure 1: The derivation of λ_E in §B.2.