LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-403659

ASP Performance Assessment: toward a science-based understanding

Ken Sale

May 8, 2008

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

ASP Performance Assessment: toward a science-based understanding

Several approaches to ASP performance can be contemplated. Perhaps the ideal would be a full cost/benefit analysis (which is probably utterly infeasible). Another approach would be a test-based figure-of-merit (FOM), this approach has the virtue of being quantitative and the challenge that each customer and application would be characterized by a different FOM. The alternative proposed here is an approach that uses information about the limits of detection of real instruments to support informed judgments.

An ideal calculation of the benefit

An ideal performance FOM would take into account, in detail, all the benefits a customer could expect to reap from the use of a system. Here benefit is defined as accomplishing the goal of interdicting a threat with a limited impact on commerce. This would mean including, for example, the entire catalog of threat objects that might be encountered along with the expected benefit (cost of failure-to-detect avoided) and the rate of attempts to deliver each object. The information for such a complete assessment will probably never be available. Nevertheless, formulating the approach may help guide a practical approach.

The expected cost of a failure to detect an item labeled i is C_i. The expected delivery attempt rate (attempts per year to deliver an object of type i per ASP at a location where ASP units are operating) is R_i. For a detection probability of P_i at each (assumed identical for convenience) ASP the benefit of the ASP would be

$$\sum_i R_i C_i P_i$$

The benefit of an entire system can be expressed in an analogous way by taking an appropriate sum over system components. It should be noted that the values of R_i are not constants of nature, they will vary over time and will depend on perceptions of the deployed interdiction system. Some have argued that as soon as P_i becomes significant (or is perceived to be so) R_i will drop drastically as the ASP-protected entry paths are effectively sealed off.

Note that only the detection probability P_i is susceptible to a purely technical assessment, and that even that assessment is non-trivial. At the present time there is no consensus on the costs due to the successful delivery of an object or estimate of the delivery attempt rate for any object. The conceptually straight-forward analysis outlined above does not take into account important considerations including the deterrent effect of a system (lowering the R_i values).

The object detection probability (P_i) can be assessed (for some given ASP hardware and algorithm and concept of operations at a specific time and place) by statistically sampling the spectrum data (incorporating the effects of shielding, masking and background) and running the system data analysis algorithm. There is no rigorous,

simple method of estimating the detection probability for a spectroscopic system using a sophisticated algorithm as there is for gross counting systems.

So much for a description of how performance ought to be calculated, consider now what can actually be done to assess performance.

A feasible approach to informing decisions

A practical approach to understanding the benefits of a system is to undertake well designed sets of tests on that system. The testing can involve measured data, synthetic data and combinations of the two.

For any kind of controlled test the results must somehow reflect a realistic concept of operations and help decision makers understand the trade-offs available among systems. One approach to doing this is to conduct tests that highlight how well the systems disentangle the physics of the problem. This approach would include considerations of how well interferences (NORM or masking) are disentangled, the ratio of net peak area to continuum or background (signal to noise), how well the effect of shielding can be teased out of the data etc. A fundamental physics- and information-based approach to performance can be used to understand the ultimate performance of a system and can transcend the details of a current algorithm that will be superseded in subsequent turns of spiral development.

Testing should include:

1) Test configurations that pose a sufficiently challenging detection problem to provide confidence that successful identification/discrimination in testing will indicate reliable performance in the field[1]
2) Tests of how well the system can disentangle threat spectrum features when (intentional, devious) masking sources are present
3) Tests of how well the system can disentangle threat spectrum features from strong (large, high activity) NORM sources
4) Tests of how well the system can extract dependable activity/mass estimates when the threat is highly shielded or shielded and masked
5) Tests of how well the system reports the higher energy part of the spectrum (beyond 3 MeV) which can be used to determine the presence of a neutron source, including an α-n source
6) Tests of how stringent a limit the system can place on the maximum activity of some isotope that might be there
7) Tests of how reliably the system can determine the presence of NORM only (no threat)

A substantial part of these fundamental performance characteristics can be reasonably well summed up in terms of signal to noise: how well does a spectral peak show up above the background and continuum. An understanding of this aspect of a system can be used to understand (in part) the potential performance of a system independent of the details of the analysis algorithm.

---

[1] Adapted from "Evaluation methodologies for active interrogation systems to detect well shielded SNM", D. Slaughter, LLNL UCRL-TR-233693

Features of the data analysis algorithm that should be tested include:

1) Does the algorithm behave well on data from an entirely new/unexpected source? I.e. does it present an error/unknown indication or a wrong indication.
2) How well does the algorithm cope with changing backgrounds and with the background depression that can be caused by an object entering the field of view?

The features outlined above can be investigated by ensuring that the test configurations include:

- Moderately strong sources behind thick, low-Z shielding. This type of configuration will provide a large continuum in the spectrum. This is the type of configuration that will easily set off a PVT system and may challange an isotope identification instrument.
- Intentionally chosen masking sources with and without the corresponding threat items. For example ^{108m}Ag and ^{177m}Lu have peaks that are close (at low energy resolution) to features in a Pu spectrum and could be incorporated into a tough masking configuration. Other combinations of isotopes can provide peaks close to other features in SNM signatures. The task of designing a set of tough cases is straight forward in concept, though it is not trivial to develop a set that can meet the constraints of reasonable cost, half-life and availability.

Good system performance against the tough configurations will provide confidence in field performance. The feedback from such testing can also provide valuable input for further research and development in support of improvements in future spirals.

A proposed method for designing physics-based testing

1) Select a few threat sources
2) Select masking sources that are
    a. Not threats
    b. Occur with some frequency in the wild
    c. Have features (lines, escape peaks, Compton edges etc.) that are close to peaks in the spectrum from the threat. The exact criteria for "close" depend on the detector technology
    d. Have manageable activities
3) Design shielding (possibly separate for threat and masking sources) to
    a. Maximize the challenge provided by masking
    b. Add up to a "reasonable" package (with respect to total weight, size, activity etc.)
4) Provide enough measurement time to get high statistics data that can be down-sampled to simulate real-time portal passages in post-measurement algorithm exercises.

The core design tasks (2 and 3 above) can be accomplished based on the existing experience base of analysts and simulations to tune up the test configurations.

Executing these steps would provide an informative data set that would enable an understanding of system performance and trade offs.

An example of a tough case is illustrated in Fig. 1. This shows a (NaI detector) Pu spectrum along with a spectrum from a combination of two isotopes that are available commercially. The similarity of the spectra gives rise to the challenge.

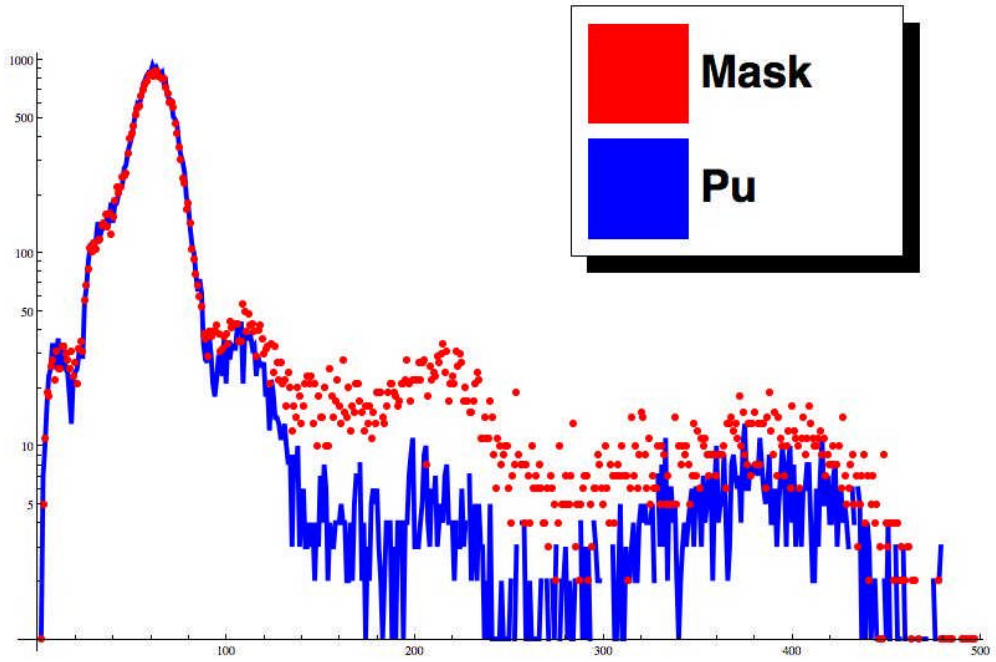

Figure 1 NaI(Tl) spectra (counts vs. energy in keV) of Pu and a two-component masking source. Poisson sampling noise is included.

Appendix A

The value of resolution

Simple statistical considerations indicate that the required counting time to establish a fixed statistical confidence for a single peak is linear with detector energy resolution: $t=q(b\Delta E+k)^2$. Thus the detector resolution directly translates to isotope identification time: longer times for worse resolution (larger ΔE). The values of the constants *q, b* and *k* are dependant on the details of the system and environment. The required time-resolution relationship is illustrated in Fig A1. Alternatively the signal to noise ratio can be used as one figure of merit for a system. Under the assumption of

[2] Detector Resolution Study, Karl Einar Nelson, LLNL, UCRL-PRES-227478

homogeneous cargo and ignoring the energy dependence of detector efficiency the signal to noise ratio can be written as[3]

$$\frac{S}{N} \propto \frac{e^{-\mu l}}{R^2} \sqrt{\frac{A \tau f}{b}}$$

where f is the resolving power ($1/\Delta E$, large values of f correspond to high resolution) of the detector system, τ is the counting time, μ the linear attenuation coefficient of the cargo, l the length through the cargo between the source and detector, R the distance from the source to the detector, A the area of the detector and b the background spectrum density (counts per second per keV). The form of the expression for signal to noise indicates that detector resolving power, detector area and counting time contribute in equal proportions to detection.

Information content

A measure of the usefulness of a system is the information content of a spectrum (roughly, the number if independent parameters that can be extracted). Realistic values for a range of detector resolutions are listed in Table A1and shown in Fig A1. The data for the table and figure are from Ref [2]

Resolution (FWHM)	Information content (theoretical max)	Information content (practical max)
0.25%	>2100	287
0.5%	1219	267
1%	645	219
2%	330	161
4%	169	62
8%	86	

Table A1. Theoretical and practical upper limits of spectrum information content as functions of detector resolution.

[3] Adapted from Appendix 10 of "Independent Review of the Department of Homeland Security Domestic Nuclear Detection Office Advanced Spectroscopic Portal Final Report" of Feb. 20, 2008

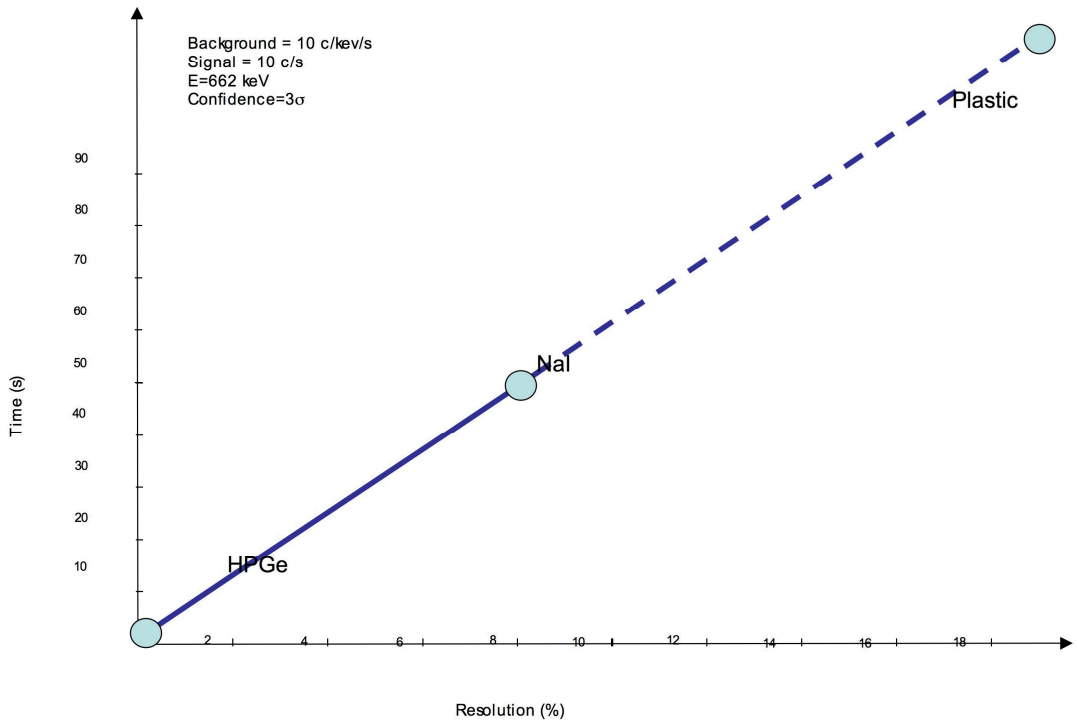

Figure A 1 Effect of detector resolution on required counting time

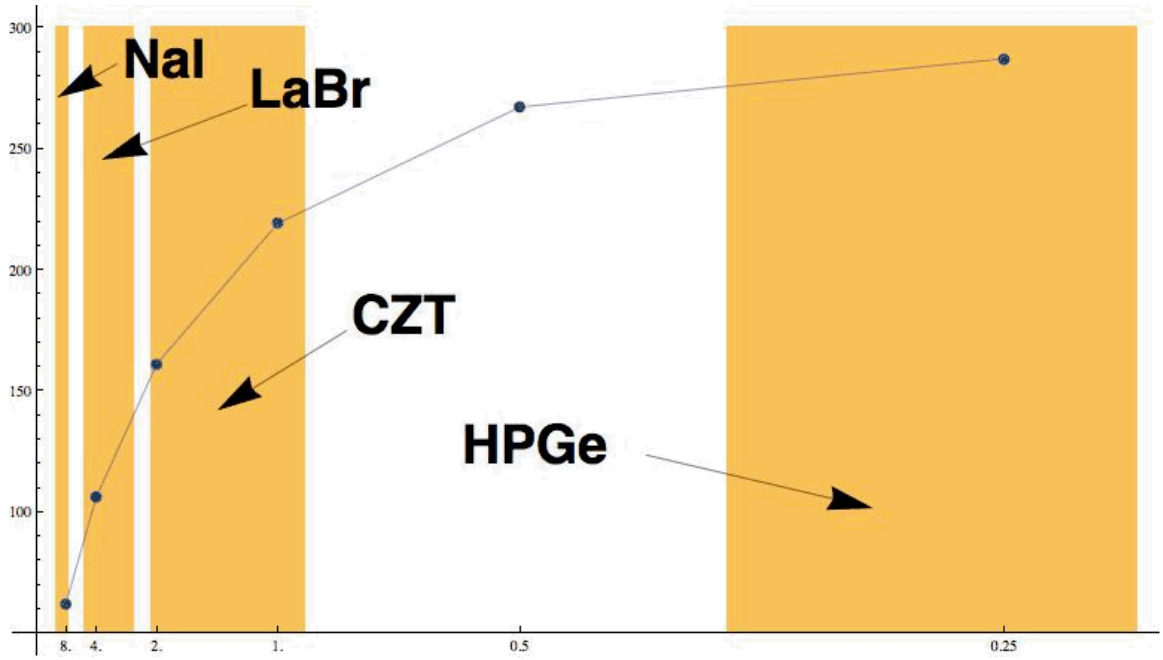

Figure A 2. Practical information content of a spectrum vs. detector system energy resolution. The four resolution bands (left to right) correspond to NaI(Tl), LaBr, CZT and HPGe.

Illustration of practical consequences

A key attribute of an ASP system is the probability of misidentification, declaring harmless material to be SNM, or declaring SNM to be non-threatening. In Table A2 ideal probability of misidentification is shown for several source/shielding configurations versus detector resolution.

Source/Shield	0.025%	0.05%	1%	2%	4%	8%
^{192}Ir 10μCi Z=5, ρr=50 g/cm^2	0.001	0.001	0.003	0.006	0.009	0.009
^{133}Ba 10μCi Z=10, ρr=30 g/cm^2	<0.0001	<0.0001	0.0002	0.0002	0.005	0.4
^{131}I 10μCi Z=10, ρr=30 g/cm^2	<0.0001	<0.0001	<0.0001	0.0001	0.02	0.3
^{67}Ga 10μCi Z=10, ρr=30 g/cm^2	<0.0001	0.0001	0.0003	0.02	0.02	0.1
1kg WG Pu 10μCi Z=25, ρr=50 g/cm^2	<0.0001	<0.0001	0.0003	0.04	0.3	0.4
^{237}Np 1g Z=10, ρr=20 g/cm^2	<0.0001	<0.0001	<0.0001	0.001	0.2	0.4

Table A 1. Idealistic misidentification probability of nuisance and threat sources as a function of detector resolution. Each spectrum had 1000 counts in it. Nuisance sources were chosen based on currently most common false alarms. Data are from ref. [2].

Appendix B

Ingredients for challenging cases

1. Shielding materials (various Z and thickness) for each source
2. Isotope sources that match SNM features and are reasonably available and have reasonable half-lives. The sources will generally have activities in the range of μCi
3. A set of specific threat item surrogates to imitate or mask.

Given these ingredients informative, tough test cases can be assembled. The ASP outputs from these tests can be used to inform further decisions on R&D or procurements.