IDA

Analysis of Potential Technologies for Air Cargo Screening: A Progress Report

Michael Finnin, Shelley Cazares, Isaac Chappell

Institute for Defense Analyses

Advanced Development for Security Applications Workshop (ADSA11)

Boston, MA

November 5, 2014

IDA | Overview of IDA

- **What is IDA?**
  - IDA runs Federally Funded Research and Development Centers (FFRDCs) for several national security agencies
  - IDA is a non-profit entity sponsored and funded by the government to provide independent, objective analyses
  - IDA does not work for or compete with for-profit entities
- **IDA Staff**
  - Research staff consists mainly of PhD-level scientists and former military
  - Expertise in a wide variety of science and technology (S&T) areas
  - Science and Technology Division (STD) performs many technology assessment functions for government S&T funding agencies such as DARPA, DTRA, DHS S&T, OUSD(AT&L), etc.
- **IDA operates 3 FFRDCs**
  - SAC (Systems & Analysis Center) supports DoD Office of the Secretary of Defense
  - STPI (Science and Technology Institute) supports the White House Office of Science and Technology Policy (OSTP)
  - CCC (Center for Communications and Computing) supports the NSA
  - IDA also operates the SAFETY Act for DHS S&T

IDA | Overview of Study for DHS S&T EXD

- **Air Cargo Metastudy Project**
  - IDA will collect and review existing test reports and studies on cargo screening technology to assess how well the technology performs against various containers, packaging (substrate), and commodities (content of cargo)
  - ***Focuses on TSA needs and any gaps in technology*** used for air cargo screening that might exist in the current state-of-the-art
  - For each technology (or technology group) and specific system within that technology, IDA will assess:
    - Are there performance metrics?
    - Are they appropriate?
    - Are there detection gaps?
    - Are follow-on studies needed?
  - Sources of Studies
    - Transportation Security Laboratory (TSL)
    - National Labs
    - JHU-APL
    - DHS S&T
    - Others as we discover them…..

IDA | Air Cargo Metastudy - Methodology

- Many existing studies and tests of technology exist can be applied to air cargo screening
- Meta-analysis attempts to provide a rigorous statistical framework in order to combine and compare the results of disparate studies.
- Key statistic - Effect Size
  - Effect size metrics may include Pd, Pfa, or other performance metrics.
  - The correlation between multiple effect size metrics must be considered.
  - Effect size metrics can be weighted for:
    - Among-study heterogeneity
    - Variance
    - Sample size
  - Moderator variables may influence effect size metrics.
- Meta-analysis should evaluate the effect size metrics across many studies including:
  - Technology
  - Packaging (container, substrate)
  - Cargo contents (Commodity)
  - Threat

Choice of effect size metrics is an important consideration in this study

IDA | Air Cargo Screening - Possible Technical Solutions

- Many potential technologies exists that can be applied to the air cargo problem
  - X-ray backscatter or thermal neutron capture
  - Photon or neutron interrogation that attempts to measure elemental composition signatures to discern threat from non-threat
  - Signatures based on capture or scattering of photon or particle
    - Examples: Nuclear Resonance Fluorescence, Fast Neutron Analysis, Pulsed Fast Neutron Transmission Spectroscopy.
- What current and nascent technology could be applied to the air cargo screening problem?
  - Depends how you want to screen air cargo (CONOPS)
    - Screen as Break-bulk or Bulk (pallet, UDL)?
    - Use the technology for initial screening, resolve a false positive, identification, etc.

Evaluating Technology with Appropriate Performance Metrics is Crucial to Develop an Effective Screening Capability

IDA | Performance Metrics - From the Scientist's Perspective

Total # of Items = 10,100		Notional System	
		"Threat"	"Non Threat"
Ground Truth	Threat	TP = 90	FN = 10
	Non Threat	FP = 500	TN = 9500

Pd = 90 / (90+10) = 90%
(near 1 is better, near 0 is worse)

Pfa = 500 / (500+9500) = 5%
(near 0 is better, near 1 is worse)

From a ***scientist's perspective***, this notional system exhibits ***excellent performance***:

- Pd = 90% → 90% of all true threats correctly cause a "threat" alarm
- Pfa = 5% → Only 5% of all true ***non***-threats ***in***correctly cause a "threat" alarm

IDA | Performance Metrics - From the Operator's Perspective

Total # of Items = 10,100		Notional System	
		"Threat"	"Non Threat"
Ground Truth	Threat	TP = 90	FN = 10
	Non Threat	FP = 500	TN = 9500

Positive Predictive Value:
PPV = 90 / (90+500) = 15%
(near 1 is better, near 0 is worse)

Negative Predictive Value:
NPV = 9500 / (9500+10) ≈ 100%
(near 1 is better, near 0 is worse)

From an ***operator's perspective***, the very same notional system exhibits ***poor performance***:

- NPV ≈ 100% → Approximately all "non threats" (i.e., absences of alarm) turn out to be truly no threat → In the absence of an alarm, the operator can rest assured there's no threat 
- PPV = 15% → Only 15% of "threat" alarms turn out to be true threats → When an alarm sounds, the operator cannot trust that there is a threat, since the system cries wolf so often 

IDA | Performance Metrics - When the Threat Prevalence is Low

Total # of Items = 10,100		Notional System	
		"Threat"	"No Threat"
Ground Truth	Threat	TP = 90	FN = 10
	No Threat	FP = 500	TN = 9500

Pd = 90 / (90+10) = 90%
(near 1 is better, near 0 is worse)

Pfa = 500 / (500+9500) = 5%
(near 0 is better, near 1 is worse)

Positive Predictive Value:
PPV = 90 / (90+500) = 15%
(near 1 is better, near 0 is worse)

Negative Predictive Value:
NPV = 9500 / (9500+10) ≈ 100%
(near 1 is better, near 0 is worse)

Prevalence = (90+10) / (90+10+500+9500) = 1%

The dichotomy between the scientist's vs. operator's perspective often emerges when the threat prevalence is very low (or very high).

When the threat prevalence is very low, most alarms are false.

Performance Metrics for Tiered Systems

Total # of Items = 10,100		Notional System	
		"Threat"	"Non Threat"
Ground Truth	Threat	TP = 90	FN = 10
	Non Threat	FP = 500	TN = 9500

Prevalence = 1%

Pd = 90%

Pfa = 5%

PPV = 15% **NPV ≈ 100%**

- From the operator's perspective, our notional system exhibits poor performance (low PPV) when used to differentiate "non threats" vs. "threats"

IDA | Performance Metrics for Tiered Systems

Total # of Items = 10,100		Notional System1	
		"*Maybe* Threat"	"Non Threat"
Ground Truth	Threat	TP1 = 90	FN1 = 10
	Non Threat	FP1 = 500	TN1=9500

Prevalence1 = 1%

Pd1 = 90%

Pfa1 = 5%

NPV1 ≈ 100%

Total # of Items = 590		Notional System2	
		"Threat"	"Non Threat"
Ground Truth	Threat	TP2 = 88	FN2 = 2
	Non Threat	FP2 = 25	TN2 = 475

Prevalence2 = 15%

Pd2 = 98%

Pfa2 = 5%

PPV2 = 78% **NPV2 ≈ 100%**

- From the operator's perspective, our notional system exhibits poor performance (low PPV) when used to differentiate "non threats" vs. "threats"
- However, this same notional system could be used as the 1st tier of a ***tiered system***, screening out the "non threats" from the "***maybe*** threats"
  - The 1st tier's low PPV does not matter*, since all "maybe threats" from the 1st tier would be further assessed in the 2nd tier
  - All that matters for the 1st tier is its high NPV, since the "no threats" from the 1st tier would not get the opportunity to be assessed further in the 2nd tier

* Provided that the 2nd tier can assess the 1st tier's 590 "maybe threats" relatively quickly & inexpensively

IDA | Performance Metrics for Tiered Systems

Total # of Items = 10,100		Notional System1	
		"*Maybe* Threat"	"Non Threat"
Ground Truth	Threat	TP1 = 90	FN1 = 10
	Non Threat	FP1 = 500	TN1=9500

Prevalence1 = 1%

Pd1 = 90%

Pfa1 = 5%

NPV1 ≈ 100%

Total # of Items = 590		Notional System2	
		"Threat"	"Non Threat"
Ground Truth	Threat	TP2 = 88	FN2 = 2
	Non Threat	FP2 = 25	TN2 = 475

Prevalence2 = 15%

Pd2 = 98%

Pfa2 = 5%

PPV2 = 78% **NPV2 ≈ 100%**

Overall System:

- **Pd** = 88 / (88+2+10) = **88%**
- **Pfa** = 25 / (25+475+9500) ≈ **0%**
- **NPV** = (9500+475) / (9500+475+10+2) ≈ **100%**
- **PPV** = 88 / (88+25) = **78%**

- From the operator's perspective, our notional system exhibits poor performance (low PPV) when used to differentiate "no threats" vs. "threats"
- However, this same notional system could be used as the 1st tier of a ***tiered system***, screening out the "no threats" from the "***maybe*** threats"
  - The 1st tier's low PPV does not matter*, since all "maybe threats" from the 1st tier would be further assessed in the 2nd tier
  - All that matters for the 1st tier is its high NPV, since the "no threats" from the 1st tier would not get the opportunity to be assessed further in the 2nd tier
- The overall system must be assessed based on ***all*** TP, FN, FP, and TN counts that did not pass to a subsequent tier

* Provided that the 2nd tier can assess the 1st tier's 590 "maybe threats" relatively quickly & inexpensively

IDA | Performance Metrics - A Summary

- Performance metrics must be carefully selected for our meta-analysis.
  - Pd and Pfa:
    - Reflect the ***scientist's perspective***
    - Are not influenced by threat prevalence
  - PPV and NPV:
    - Reflect the ***operator's perspective***
    - Are influenced by threat prevalence
- The appropriate performance metrics depend on how the system will be used, particularly if the system is only ***one tier of an overall system***
  - ***PPV*** does not matter for the 1st tier (screener), provided that the 2nd tier can operate relatively quickly and inexpensively
    - Keep in mind that the 2nd tier will operate on fewer items than the 1st tier
  - ***NPV*** is the most important metric for the 1st tier (screening tier)
  - The metrics selected for the overall system must include ***all*** TP, FN, FP, and TN counts that did not pass through a subsequent tier
- Other performance metrics may further reflect the operator's perspective, such as ***throughput***, ***workload***, ***usability***, etc.
  - Very few reports have data that could be used to estimate these metrics.

IDA | Testing Documents

- Documents describing tests of different technologies, commodities, and packaging:
  - Cover a testing/reporting period of 1998 - 2013
  - Test procedures, metrics, and types of commodities vary widely among reports
  - Some reports are detailed tests while others are “Quick Looks”
  - Many reports are associated with separate test plans
  - Threats are coded in many later reports for classification reasons
- Technologies Considered
  - Explosive Trace Detectors (ETDs)
    - Trace detection based on chemical signature
    - Examples: Ion mobility spectroscopy or infrared spectroscopy
  - Explosive Detection Systems (EDS)
    - Radiation beam (photons-X-ray, gamma or particles-neutrons) interrogates sample
    - Signatures based on shape, density, or elemental composition of sample

IDA | Testing Documents - Building Matrices

Matrix Axes

- Technology
- Commodities
  - Types and number vary
  - "Standard 8" →

EE	Electronic Equipment
WA	Wearing Apparel
PM	Printed Matter
MP	Machine Parts
MDG	Miscellaneous Durable Goods
FF	Fresh Flowers
PR	Fresh Produce
SM	Seafood & Meats

- Packaging
  - Break-bulk
  - Containerized
  - Palletized
- Substrate
  - ETD specific variable
  - Represents the sampling surface the ETD encounters →

Plywood
Cardboard
Packing Tape
Stretch Wrap
ABS Plastic

- Threats
  - Coded in later reports
  - Quantities expressed in undefined "threat weights"
- Performance Metrics
  - Scientist's Perspective
  - Operator's Perspective

IDA | Air Cargo Commodities

Code	Category Description	Typical Commodities
CHEM	Chemicals	Chemicals, alcoholic beverages, glass, chemical and fuel oils, pharmaceuticals
EE	Electronics	Electronic components, computer, medical and lab equipment
FF	Fresh Flowers	Flowers and herbs
HR	Human Remains	Human remains, organs and blood products
LA	Live Animals	Pets, tropical fish, live animals for restaurants
MDG	Miscellaneous Durable Goods	Non-metallic mineral products, base & construction material, furnishings, misc. manufactured products
MP	Machine Parts	Machinery & vehicle parts
MULT	Multiple	Mixed commodities (UDLs)
PM	Printed Materials	Newsprint, magazines, books
PP	Paper Products	Non-printed paper, plastic & rubber products
PR	Produce	Fresh produce, grains & animal feed, perishables, bakery & dried foods
SM	Seafood and meats	Fresh & frozen seafood & meat products
WA	Wearing apparel	Clothing
UNK	Unknown	No commodity info

IDA | Potential Matrix: ETDs Versus Substrate

Technology	Plywood	Cardboard	Packing Tape	Stretch Wrap	ABS Plastic
GE/Morpho Itemiser 2					
GE/Morpho Itemiser DX					
Smiths Ion Scan 400B					
Smiths Ion Scan 500DT					
Smiths Sabre 4000 (vapor)					
GE/Morpho Mobile Trace					
GE/Morpho Hardened Mobile Trace					
Implant Sciences ACSS QS-H300					
Fido XT					
Fido Scout					

IDA | Issues and Interim Findings

- Air cargo environment provides unique challenges for explosives detection
  - Packaging and commodities are highly varied
  - Very little standardization or predictability on how goods are shipped on passenger aircraft.
- No technologies are specifically designed for air cargo screening
  - Technologies are repurposed and optimized for other environments
  - Air cargo has unique technology requirements in density, size, and packaging
  - Air cargo has unique CONOPS requirements
- Testing documents
  - Over 15 years of testing with variable:
    - Protocols
    - Personnel conducting tests
    - Testing goals
    - Metrics for success - if at all
- IDA quick analysis
  - Currently deployed technologies may have a specific role in a multi-tiered screening system which would depend upon their particular performance metrics.
  - Number of technology tiers that would be required depends on how well individual technologies perform to resolve the “maybe threats” issue (see slides 7-13)

Contact Information

Michael Finnin
Institute for Defense Analyses
Science and Technology Division
4850 Mark Center Drive
Alexandria VA 22311

(703) 578-2737
mfinnin@ida.org

Thanks to Program Managers Stephen Surko and David Throckmorton and the Department of Homeland Security Science and Technology Directorate for sponsoring this work.

IDA | Air Cargo Screening - Background

- Passenger aircraft transport is “belly-loaded” with cargo for increased revenue
  - Originating and Trans-shipped
    - Originating cargo is cargo that is initially delivered to the air cargo facility
    - Trans-shipped cargo originates at one facility and passes through another facility
  - Exempt and Non-exempt
    - TSA has established rules for cargo that must be screened and cargo that is exempt from screening
- Packaging
  - Containerized
    - Cargo arriving as a bulk shipment in a Unit Load Device (ULD)
  - Palletized
    - Bulk shipment wrapped in plastic on pallets
  - Loose Cargo
    - Individual pieces
    - Can be result of breaking above bulk shipments - “Break-bulk”
- Commodities - Contents of air cargo

Packaging, threat, and commodity type influence the choice of screening procedure and technology employed.

IDA | Air Cargo Screening vs Baggage Screening

- Passenger bag screening technologies may be applicable to air cargo
  - Mature and deployed at majority of airports
- Differences between air cargo and passenger baggage
  - Pallets and ULDs have different physical characteristics
    - Much larger internal space to be interrogated for air cargo
  - Contents of these packages (commercial commodities vs personal effects)
    - Contents of air cargo vary significantly across the Enterprise
  - Traditional baggage screening systems are not engineered to accommodate air cargo screening

Unit Load Device

Typical Dimensions:
160 x 220 x 320cm^3
Up to 1600kg
Average Density 23 g/cm^3

Suitcase

Typical Dimensions:
50 x 50 x 100cm^3
Up to 25kg
Average Density 5 g/cm^3