Informational Sensing for Nonproliferation[1]

James Kornell,* Zoe N. Gastelum†, Bethany L. Goldblum‡

*Special Technologies Laboratory (NSTec), Santa Barbara, California 93111, USA, korneljm@nv.doe.gov
†Sandia National Laboratories, Albuquerque, NM, USA
‡Nuclear Science and Security Consortium, University of California, Berkeley, California 94704, USA

MOTIVATION

Verifying nuclear nonproliferation is difficult, in that the goal is to continually confirm the absence of undeclared nuclear materials and activities, to say with high confidence that what we don't see doesn't exist. The nuclear security community has largely relied on physical sensing for detection of potential clandestine activity. There are many kinds of information that could complement physical sensing, including open source data, publicly available images produced by commercial or scientific satellites, networked sensors, various societal verification tools, quantitative political science datasets, and social media.

The potential contribution from open-source data is critical to improving verification and detection. For example, modern smartphones, of which there are expected to be five billion by 2020, can without or with inexpensive modification sense infrasound, sound, infrared, visible, and hyperspectral light, gamma radiation, and seismic activity. For most of these, there are analytics companies and interfaces that provide potentially useful content. Both Google and Microsoft provide open software that could, with training, tell us where every truck carrying a 48Y cylinder was seen in the background of a Facebook or Instagram post for any given day across the entirety of Canada or Australia. With so many types and such volume of data from so many disparate sources, the nonproliferation community needs effective methods for systematic analysis of these data and an approach to understanding their interdependencies.

INFORMATIONAL SENSING

Informational sensing comprises *techniques for identification of informational signals in open source and other data*. These techniques derive from complexity science and include network analysis, game theory, nonlinear dynamics, pattern recognition, machine learning, and artificial intelligence. The dynamic combination of physical and informational sensing may allow significantly improved proliferation detection.

A concrete example from a different domain may make this more tangible. Some Western US farms control watering by combining physical sensing of soil moisture, temperature, and direct-sun exposure, with informational sensing for location in the field (fields are not perfectly flat), predicted weather, and historical rainfall patterns. The complement of physical and informational sensing is much more effective than either could be on its own.

Advances in information technology have yielded a radical democratization of content and access to the tools to distribute it. This provides both opportunities and challenges for nuclear nonproliferation. The combination of physical and informational sensing enables previously impossible monitoring and verification potential. Yet, it also provides access to powerful capabilities to those with bad intentions.

It is difficult to directly detect physical signal when skilled people want to hide it. Adding informational signals and signatures to physical sensing promises to aid analysts in allocating their focus of attention to the most productive places, times, and circumstances.

COMPLEX SYSTEMS, STRONG CONSTRAINTS

Development of informational sensors faces different problems than the corresponding development of physical sensors. For the latter, a well-understood signal is the target, and the significance of detection is left to the user of the sensor. Informational signals are context-sensitive, though: high gamma is (if accurately measured) a unique signal, while a small, dense packages moved in the depth of the night is ambiguous without informational context. Since informational signals are likely to be individually weak, only by correctly aggregating them will useful mutual support with physical sensing be provided. Correct aggregation, though, is far from trivial. Signals can participate in multiple networks, and have different meanings in each. For instance, hydrogen fluoride (HF) is used in both the conversion and the reprocessing stages of the nuclear fuel cycle (NFC). An indirect effect of HF use might be corrosion of the paint on cars that regularly park next to a conversion or reprocessing facility. However, as a metals solvent, HF is used in many other industrial processes, so detection of paint corrosion without supporting context may not be useful. Table I shows sample indirect evidence of activity and the signals that might be found in open data.

Much of the context needed to usefully correlate informational signals and signatures with physical sensing comes from the intrinsic properties of the NFC. The physics, chemistry, industrial processes, and supply chain, the necessary timing of processes, the characteristic co-location of some processes and even stages, can all be used to localize and constrain potential activities of interest by mapping indirect evidence to underlying physical and chemical requirements.

---

[1]This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Signal	*Seismic* Public sources, smartphones	*Infrasound* Inverse-pressure doors	*Neutron* Darkened glass in photos	*Gamma* Smartphones, specialized equipment	*Effluents* Animal behavior	*Incongruous chemicals* Labeled actinide solvents in unmarked facility
Signal	*Heat* Exposed water warmer than background	*Mass* Vehicle loads, off-vehicle movement	*Density* Manner of movement of small packages	*Ionization* Unlikely from SNM, may indicate VOCs	*Cell damage* Foliage, hospitalization, metabolites	*Corrosion* Visible on cars, buildings, concrete, tire sidewalls

TABLE I. Sample direct and derived indicators of nuclear/radiological materials and chemicals

For example, a fully-loaded 48Y cylinder weighs roughly 8,450 kg and can only travel by major road or rail routes. There are limits on how fast it can travel, and expected rates of progress. As we construct a framework for informational sensing, these can be correlated with evidence from open data to tightly constrain informational search for a specific shipment. Image data from social media and commercial satellites can be localized, and those identifications on the restricted set of possible roads and within the plausible physical and temporal range of the original sighting can be used for independent tracking. The rapid commercially-driven advances in semantic image and video content understanding that make this possible call for almost monthly monitoring to keep up. For example, Google's Open Image API understands 'transport' when given an image of two people carrying a detector. Applying the constraints we know about materials production and transport lets us zero in on where to look; advances in mapping, 3D reconstruction, and image semantics may make it possible to find the evidence we need.

NETWORKS, SIGNALS, INFORMATION SOURCES, AND EXAMPLES

We focus on the interlinked and interdependent networks within which the NFC exists. From a network science perspective, each stage represents a node (or at finer resolution, cluster of linked nodes) in the materials acquisition network, e.g., a small installation to separate uranium out of phosphate rock under the guise of fertilizer production would fall under 'Sourcing.' Nodes and links have attributes such as equipment, location, consumables (like ore, chemicals, and/or electricity), transportation, and so on. Each of these attributes and links has associated physical and informational signals and signatures. A set of informational and physical signals based on the particular proliferation pathway can then used as proliferation indicators.

Our networks fall into five categories: political, military, economic, infrastructure, and knowledge. There are multiple networks within each category, e.g., political networks include alliances, regional powers, religious or cultural alliances/antagonisms, and so on. Both temporal extent and content precision vary across networks. We know the steps, the chemistry and industrial processes, the heat and power requirements, and the minimum time required for converting ore into yellowcake (impure U_3O_8 into UF_6) with good precision, while political and economic factors work more slowly and with less predictability. By combining linked interdependent networks into multiplex networks,[1] two benefits result: (a) the constraints intrinsic to the physical and chemical processes can reduce the range of possibilities (search space) for the political, military, and economics; and, (b) influence can propagate in both directions, down from political decision-making and up from physicochemical requirements.

There are multiple sources, summarized in Table II. Not all sources will have equal availability, completeness, and accuracy for all nations. However, just as what were previously 'national technical means' are now multiple commercial vendors for satellite imagery, information leakage is only imperfectly controllable, even for tightly-controlled nations like China and Iran.[2]

ANALYSIS METHODS

Systematic analysis of heterogeneous networks and of networks of networks group roughly into three types: network content and semantics, network structure, and statistical analysis/machine learning.

Network content includes all of the specifics that give power to the constrained correlation approach we advocate. There are two approaches (at least), semantic networks and dynamical construction.

Network semantics, also called semantic graphs or ontologies, can structure content search by encoding known relationships, dependencies, content, and co-occurrences in a logically consistent way. For example, the six steps in wet conversion can be linked in a set sequence, with chemical, temperature, timing, and heat/cooling requirements specified. Or, a semantic graph could be used in image analysis to identify potential processing facilities by the conjunction of perimeter fences, guard stations, buildings, HVAC patterns, and so on.

Alternately, network construction can be dynamic, with constraints propagating deterministically, e.g., allowing multiple possible networks (interpretations) to be dynamically generated with data (or whole networks) being discarded when inconsistent with physical or chemical NFC processes. In the

[1] Multiplex network nodes have links to multiple networks; a reprocessing plant, for example, connects to the water, sewage, electrical, and road networks, to the equipment supply chain network, to the chemical supply network, and so on.

[2] We do not here address the ethical, legal, and social issues inevitably entangled with informational sensing. See http://thebulletin.org/potential-and-pitfalls-societal-verification for an overview of some of the perspectives.

common case of plausible but underdetermined data, similarity to prototypical situations can be measured by informational distance (Shannon entropy), *i.e.*,

$$H(X) = \sum_{i=1}^{n} P(x_i)I(x_i)$$

. . . where the entropy of NFC stage or process X is measured against the probability (P) and the actual information (I) itself. For instance, the probability that joint detection of tributyl phosphate with a volatile organic compound like kerosene or dodecane is correlated with conversion or reprocessing is high if it is detected in an isolated area, but low if detected at an airport (TBP is used in aircraft braking systems, and kerosene is used as fuel.) Informational entropy is powerful but not intuitive. It might be helpful to think of an example of 1000 pixels for a photo representing the ideal ($P = 1$) item of interest, and (assuming registration and/or rotation) the number of non-matching pixels equaling the informational distance. Both kinds of network semantics can work together, and in fact need each other.

Network structure analysis examines the connectivity structures of networks to discover relationships (*e.g.*, [1]), to detect gaps (*e.g.*, [2]), and to model interdependence (*e.g.*, [3]). Multiplex networks, as we are using, have the useful property of allowing the participating networks to be modeled using dynamics that make sense for the particular network, whether historical, Bayesian, deterministic, semantic graph, or nonlinear/nonequilibrium. Network interdependence is of particular interest because it provides a quantitative method for representing potential cascades, in which change in a local network state can have global consequences. These dependencies are rarely obvious, but can provide potential clues for heightened focus of attention.

Statistical techniques in machine learning are central to many of the content analytics within the various NFC-related networks upon which we rely. For many of the analytic content sources cited in Table II, re-entrant many-layered neural networks (deep learning) systems are central [4]. The utility of deep learning systems is directly correlated with the number and quality of available examples. Google is better than humans at facial recognition, having many millions of examples. However, deep learning would not help discriminate cooling ponds under ice in an IR image, as a human analyst can, since there are so few examples.

Our approach does not focus on the social networks that have been used, with mixed success, to understand, *e.g.*, hostile non-state actors. While such graphs could potentially be overlaid on our work, we are focused on physical and pragmatic constraints that will allow us to filter and extract useful information from the daily open media deluge. Of the ~1.8 billion photos uploaded to social media daily, only a very few will have content relevant to proliferation detection. Our choice of multiplex constraint networks is driven by the need to discard the overwhelming majority of the constantly-generated content while capturing the tiny scraps of relevant information.

LIMITS OF APPROACH

The knowledge barrier to making a basic, functional fission device has fallen [5], and the barriers to designing

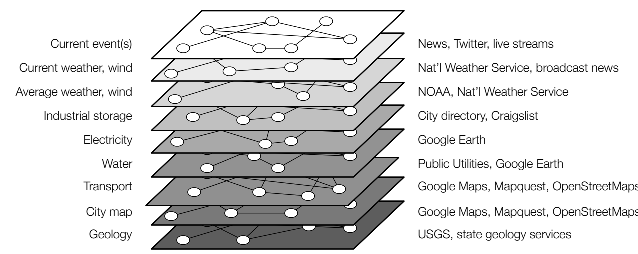

Fig. 1. *Each network contributes. In the geology network, thorium and granite mask gamma, and radon masks SNM ionization. We might note rented industrial facilities or storage with activity but to which mail is never delivered, particularly if they are located upwind of a city center.*

and manufacturing a sophisticated device will follow, if they haven't already. Non-state actors don't need the arsenals and concomitant industrial facilities that state-level actors do. Tracking materials and discovering potential proliferation activity is urgent, and the need will increase.

At the same time, the number of analysts is inadequate, the volume of content they have to sort through is more than a human can manage; and even something as simple as a travel budget is often unavailable.[3]

The approach we have described is not a superhuman solution, as are beginning to be seen in bounded domains [6]. Rather, it is intended to aid analysts in focusing on the most likely times and places to discover proliferation-relevant activities. Frank Pabian, of Los Alamos National Laboratory, has noted that a 'serendipity factor' has been necessary to the success of much previous work in open source nonproliferation monitoring [7]. The multiplex network formalization of NFC processes has the potential to greatly reduce the need for luck that has made social media more a promise than a resource. As we start to see clusters of correlated informational and physical signals, we can use the basis set to evaluate the likely proliferation network that the signals arose from and then use that to target investigation.

CLOSING REMARKS

The value and even the practical possibility of the approach described above is not yet known. There are many cases where data fusion has enabled improved capabilities, and informational sensing and analytics are a strategic resource for commercial, manufacturing, and retail industries worldwide. The space of potential covert proliferation is much larger than well-understood commercial applications, though, and even if distinct, repeatable predictive patterns exist human cognitive limits may not allow us to see them (see [8] for an example from a far simpler domain). The challenges to global peace and security posed by nuclear proliferation and terrorism continue to grow. With advances in computing and increasing data available, we as the nonproliferation community are obligated to develop new methods and innovate in this space for the promise of a better world.

[3] Personal communication, independent sources

ACKNOWLEDGMENTS

This material is based upon work supported in part by the Department of Energy National Nuclear Security Administration through the Nuclear Science and Security Consortium under Award Number(s) DENA0000979 and through DOE NNSA contract number DE-AC52-06NA25048.

REFERENCES

1. A. WIERZBICKI, U. BRANDES, F. SCHWEITZER, and D. PEDRESCHI, editors, *Advances in Network Science*, *Lecture Notes in Computer Science*, vol. 9564, Springer International Publishing, Cham (2016).
2. A. J. GATES and L. M. ROCHA, "Control of complex networks requires both structure and dynamics," *Scientific Reports*, **6**, 24456 (Apr. 2016).
3. B. GAO, B. BARZEL, and A. L. BARABASI, "Universal Resilience Patterns in Complex Networks," *Nature*, **530**, *307* (2016).
4. Y. LECUN, Y. BENGIO, and G. HINTON, "Deep learning," *Nature*, **521**, *7553*, 436–444 (May 2015).
5. R. MUNROE, *Thing Explainer: Complicated Stuff in Simple Words*, Houghton Mifflin Harcourt (2015).
6. E. DAVIS, "Ethical guidelines for a superintelligence," *Artificial Intelligence*, **220 IS -**, 121–124.
7. F. V. PABIAN, G. RENDA, R. JUNGWIRTH, L. K. KIM, E. WOLFART, and C. G. G, "Open Source Analysis in Support of Nonproliferation Monitoring and Verification Activities: Using the New Media to Derive Unknown New Information," Tech. rep. (2014).
8. R. GALLOTTI, M. A. PORTER, and M. BARTHELEMY, "Lost in transportation: Information measures and cognitive limits in multilayer navigation," *Science Advances*, **2**, *2*, e1500445–e1500445 (Feb. 2016).

Data Type	Data Source	Content Analytics	NFC stage(s)
Maps	USGS, Google Maps/Google Earth, National Geographic Maps, UN GIS		Transport, frame for all stages
News	BBC, Al Jazeera, AP, Reuters, Agence France-Presse	tldr, Agolo, Clipped	Sources, Power generation
Technical publications	arXiv.API, CrossRef REST API, IEEE Explore API		Sources (esp. seawater), Enrichment (new methods)
Trade data	UN ComTrade, World Bank, US Census Bureau		Sources, Transport, Power generation, Storage
IAEA databases	PRIS, CNPP, RRDB, ARIS, ENDF (Brookhaven), ICSRS, UDEPO		Sources, Enrichment, Fuel fabrication, Power generation, Reprocessing
Patents	EPO, Google Custom Search/Patents	Espacenet	Sources (seawater), Sources, Enrichment (new methods)
Public records	(Too numerous to list)		(All stages)
University and NGO sites	Arms Control Institute, Nuclear Threat Initiative, Bulletin of the Atomic Scientists, &c.		Sources (esp. seawater), Enrichment (new methods), Reprocessing (separation, new methods)
Nuclear trafficking data	IAEA Illicit Trafficking, NURE, US EPA Uranium Mines & Mills DB		Sources, Enrichment, Reprocessing
Blogs and microblogs	Nukes of Hazard, Arms Control Wonk, Restricted Data, Arms Control NOW		Trends and political factors
Images	YFCC100M, Google Cloud ML, Microsoft API, Hadoop HIPI	Ditto Labs, Imagga	Sources, Transport, Enrichment, Fuel fabrication, Power generation, Reprocessing, Storage
IR images	Smartphones, PIR/IoT, personal gear		Transport, Storage
Hyperspectral images	Smartphones, personal cameras, commercial overflights		Sources (unconventional), Testing
Lidar	USGS, Open Topography, Lidar Online, NEON, commercial sources		Sources, Testing
Satellite imagery	DigitalGlobe, Terra Bella, Planet Labs, Airbus Pleiades, RADARSAT		Sources, Fuel fabrication, Power generation, Storage, Transport
SAR	UNAVCO, SSARA, UrtheCast (in dvmt)		Enrichment, Fuel fabrication, Reprocessing
Seismic	SEI Seismic Exchange, JPL		Testing
Video	YouTube, Vimeo, GoPro, dashcams		Sources, Transport, Enrichment, Fuel fabrication, Power generation, Reprocessing, Storage
Audio, incl. infrasound	Google Cloud API, Microsoft API, CarCapture		Transport, Enrichment (centrifuge)
Online communities	Nuclear Issues Forum, PONI Forum, Reddit		Power generation
Microtasking	Amazon Mechanical Turk, Microtask, Fiverr, Gigwalk		Enrichment, Fuel fabrication, Power generation, Reprocessing, Storage (human search/validation)
Contests	Kaggle, InnoCentive, TopCoder, TunedIT, Challenge.gov		Enrichment, Reprocessing, Storage (all: irregularities in data)

TABLE II. Open source and social media, with analytics providers or services and relevant NFC stage(s)