More is Better: The Analytic Case for a Robust Suspicious Activity Reports Program

James E. Steiner

In his March 2009 testimony, Gregory Nojeim warned Congress of the potential danger to civil liberties posed by the government's suspicious activity report (SAR) program. But Nojeim, director of the Project on Freedom, Security, & Technology, raised another concern – that "the security 'bang per byte' of information gathered may be diminishing. While 'stove piping' was yesterday's problem, tomorrow's problem may be 'pipe clogging,' as huge amounts of information are being gathered without apparent focus."[1] In concluding his testimony, Nojeim recommends:

> The Subcommittee should test whether SAR reporting is both effective and efficient in preventing terrorism.... SAR reporting may or may not be the best way to collect the 'dots' that need to be connected to head off terrorist attacks; whether it is or is not should be tested. Because the SAR reporting system will result in the collection of so much information about innocent activities, it seems that it would be good to know at the front end that the results are likely to be worth the risks.[2]

A subsequent CRS study, in November 2009, endorsed Nojeim's suggestion questioning the need for a data-intensive program and made a similar recommendation: "Congress may be interested in how a future SAR Program Management Office intends to address this problem – specifically, which agency or agencies will be responsible for quality control of SARs [sic] to prevent system overload from irrelevant or redundant ones."[3]

This article acknowledges the progress made in protecting civil rights – an area of legitimate concern – but rejects categorically the call to reduce or limit the size of the SAR program. Two analytic requirements for the collection of more rather than less information through the SAR process are presented, to increase the probability of identifying pre-operational terrorist activity and to improve the efficiency and effectiveness of critical infrastructure protection regimes. In statistical analysis, more is better.

The SAR Program and Process

Since 2007, the U.S. homeland security, law enforcement, and intelligence communities have formally recognized the usefulness of SAR in counterterrorism. The U.S. government defines a SAR as "official documentation of observed behavior that may be indicative of intelligence gathering or pre-operational planning related to terrorism, criminal, or other illicit intention."[4] The primary impetus for the federal government's National SAR Initiative (NSI) is intelligence support to law enforcement, although officials from the critical infrastructure (CI) protection world have recognized that SAR can help them improve their performance.

Law enforcement has been collecting SAR (or something similar, such as field interviews of suspicious individuals) for decades. The NSI program endeavors to formalize a nationwide, uniform process for evaluating and recording this information and then making it available to appropriate personnel and organizations through the

Information Sharing Environment (ISE). The SAR concept and the NSI were both originally developed to provide direct support to law enforcement. Here is the typical process in brief: law enforcement intelligence analysts receive and evaluate all new SAR. In those relatively rare cases where a single SAR is both credible and actionable, the information is embedded in a short analytic report and provided directly to law enforcement for immediate counterterrorism action. These leads go to the local Joint Terrorism Task Force (JTTF), which has first right of refusal in terrorism cases. If the JTTF chooses not to follow-up, the lead is passed to state or local police for action. The figure below illustrates the Notional SAR Process in detail.[5]

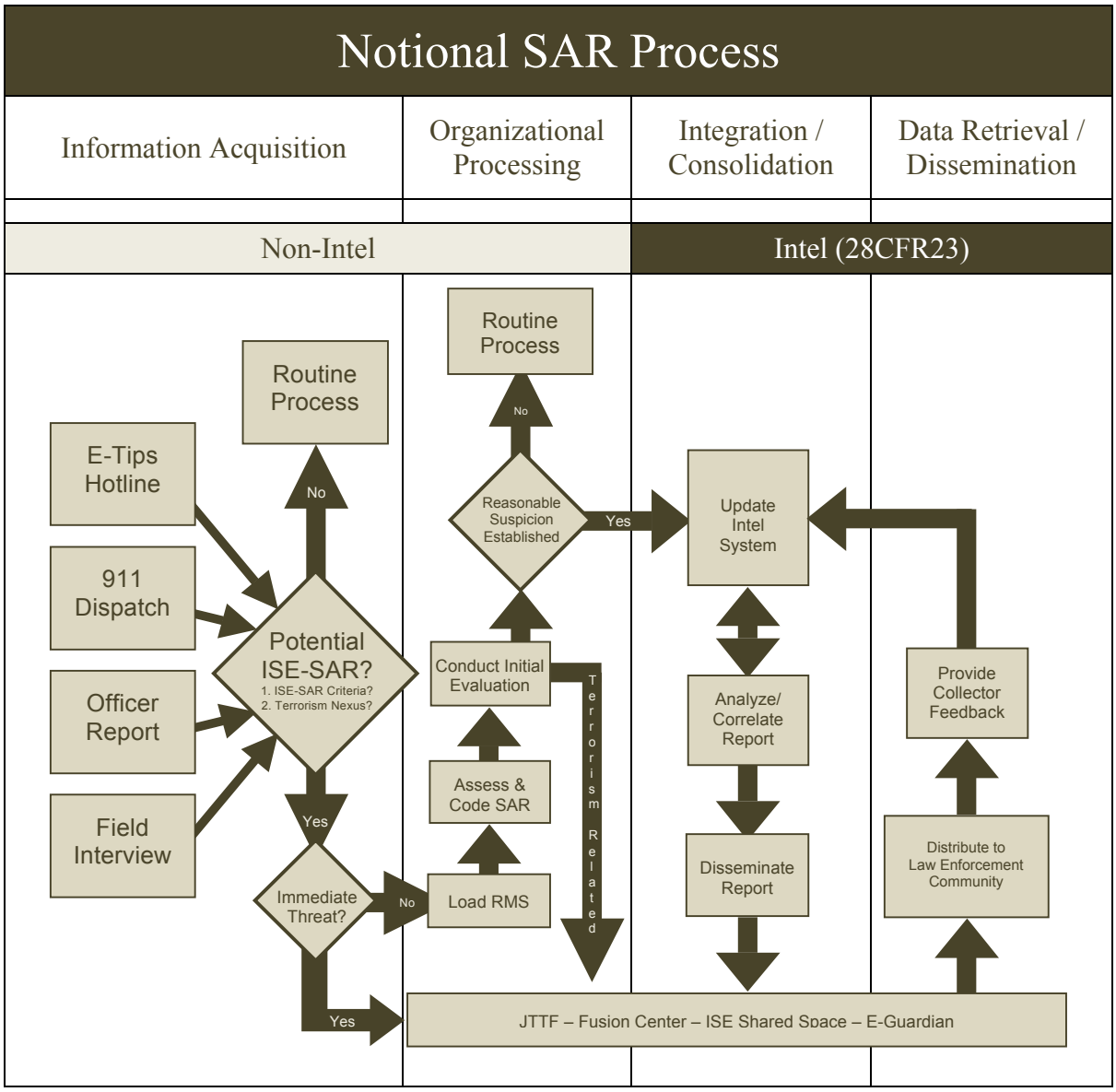

Figure 1. Notional SAR Process

This graphic has a familiar feel since it shows the traditional intelligence cycle as applied to suspicious activity reports. It begins with **collection** (information acquisition), moves to the (organizational) **processing** of that information, then on to **analysis** (integration/consolidation) and ends with final **dissemination** (and data retrieval) of the SAR and SAR-based analysis to the law enforcement customer (JTTF). Although not shown clearly, the cycle is completed through the "**feedback** to the collector" function in the last column of the graphic.

SAR originate from a wide variety of sources, including law enforcement officers, public and private sector security, and the public through "phone in" calls to hotlines, and are reported to a large number of federal, state, local, tribal, and territorial law enforcement organizations. From an intelligence perspective, the lack of direct control over some of these intelligence sources results in exceptional difficulty in assessing the quality of SAR. The NSI/ISE leadership is acutely aware of this problem as highlighted in the original "Nationwide Suspicious Activity Reporting Initiative Concept of Operations":

> A standard reporting format is a key element of the effective implementation of a SAR program. A standardized report provides a mechanism for the efficient transition of the suspicious activity from the line-level officer to the agency management. This process will ensure that the suspicious activity is being collected and reported correctly and will regulate the reporting procedures across the agency.
>
> Additionally, in order to identify local, regional, and national trends in crime and terrorist precursor activity, a common national set of data collection codes needs to be adopted to ensure seamless sharing and analysis of suspicious activity. This national standard of codes will ensure that patterns of criminal behavior are identified and handled properly. The establishment of these codes needs to be the result of evaluation and determination that the activities to be collected are likely precursors of terrorist activity."[6]

If anything, this focus on SAR quality (a necessary condition for using SAR in the analytic process) has increased as seen in the extensive treatment dedicated in the "Final Report: Information Sharing Environment – Suspicious Activity Reporting Evaluation Environment."[7]

Legitimate Civil Rights Concerns

Coming from a civil liberties perspective, Nojeim is most concerned about the range and number of collectors and the fact that despite a plethora of "quidelines" for collecting, handling, and storing SAR, the *collection guidelines "fail to provide adequate guidance"* (emphasis added).[8] This legitimate issue raised by Nojeim was directly addressed by the NSI/SAR leadership in developing their January 2010 report on the NSI-SAR Evaluation Environment. In fact, no less an organization than the ACLU has provided kudos to the NSI SAR leadership, its efforts, and its results:

> The ACLU released a report criticizing these SAR programs in July 2008. In response, ISE program manager Thomas E. McNamara and his office worked with the ACLU and other privacy and civil liberties groups, as well as the LAPD and other federal, state and local law enforcement agencies, to revise the ISE SAR functional standard to address privacy and civil liberties concerns.

> The revised ISE guidelines for suspicious activity reporting, issued in May 2009, establish that a reasonable connection to terrorism or other criminal activity is required before law enforcement officers may collect Americans' personal information and share it within the ISE.... The revised ISE functional standards also make clear that behaviors such as photography and eliciting information are protected under the First Amendment, and require additional facts and circumstances giving reason to believe the behavior is related to crime or terrorism before reporting is appropriate. These changes to the standard, which include reiterating that race, ethnicity and religion cannot be used as factors that create suspicion, give law enforcement all the authority it needs while showing greater respect for individuals' privacy and civil liberties. *We applaud the willingness of the ISE Program Manager to engage constructively with the civil liberties community and to make significant modifications to the functional standard to address the concerns presented* (emphasis added).[9]

So far, so good. The civil liberties community and the SAR leadership agree that privacy issues are being addressed through guidance on how to handle such information.

Criticism Beyond Civil Liberties

Unfortunately the ACLU also picked up and repeated Nojeim's concern over the sheer *volume* of SAR, implying that a smaller program would be better in terms of program efficiency and effectiveness. In other words, these civil rights advocates and organizations are now evaluating the analytic merits of intelligence officers having more or less data available to do their job, a subject well beyond their civil liberties expertise.

Specifically, the ACLU continues:

> This overbroad reporting mandate is not just constitutionally questionable; it's also counterproductive. *These orders, if taken seriously by local law enforcement, can yield only one outcome: an ocean of data about innocent individuals that will overwhelm the investigative resources of the authorities* (emphasis added).[10]

In fact, this final conclusion is both naïve and incorrect. Intelligence analysts actually need an "ocean of data" concerning suspicious activity (but NOT the personal identities of those engaging in such activity) to successfully apply sophisticated statistical analysis techniques that have the potential to discern between benign activity and true terrorist precursor activity, thereby reducing the investigative load. Similarly, analysis of comprehensive data on suspicious but benign behavior near critical infrastructure is the first step in developing security programs to protect that infrastructure.

Why Analysts Need a Massive SAR Database

Figure 2 lays out a simplified representation of intelligence-led SAR support to law enforcement. As shown by the two arrows leading to law enforcement actions, there are two ways SAR-based intelligence reaches police authorities. The direct approach was discussed above and is straightforward. In the rare cases where a SAR is a clear indicator of threat, the SAR information moves quickly and directly from the intelligence realm to law enforcement for investigation.

The "indirect" method is fundamentally distinct from the direct method. In the indirect case, each new SAR is processed and evaluated by both intelligence and critical infrastructure analysts to determine whether there are *non-obvious* reasons for law

enforcement follow-up. To accomplish this task, the analysts must determine whether the new SAR is an anomaly when compared to the results of "pattern analysis models."

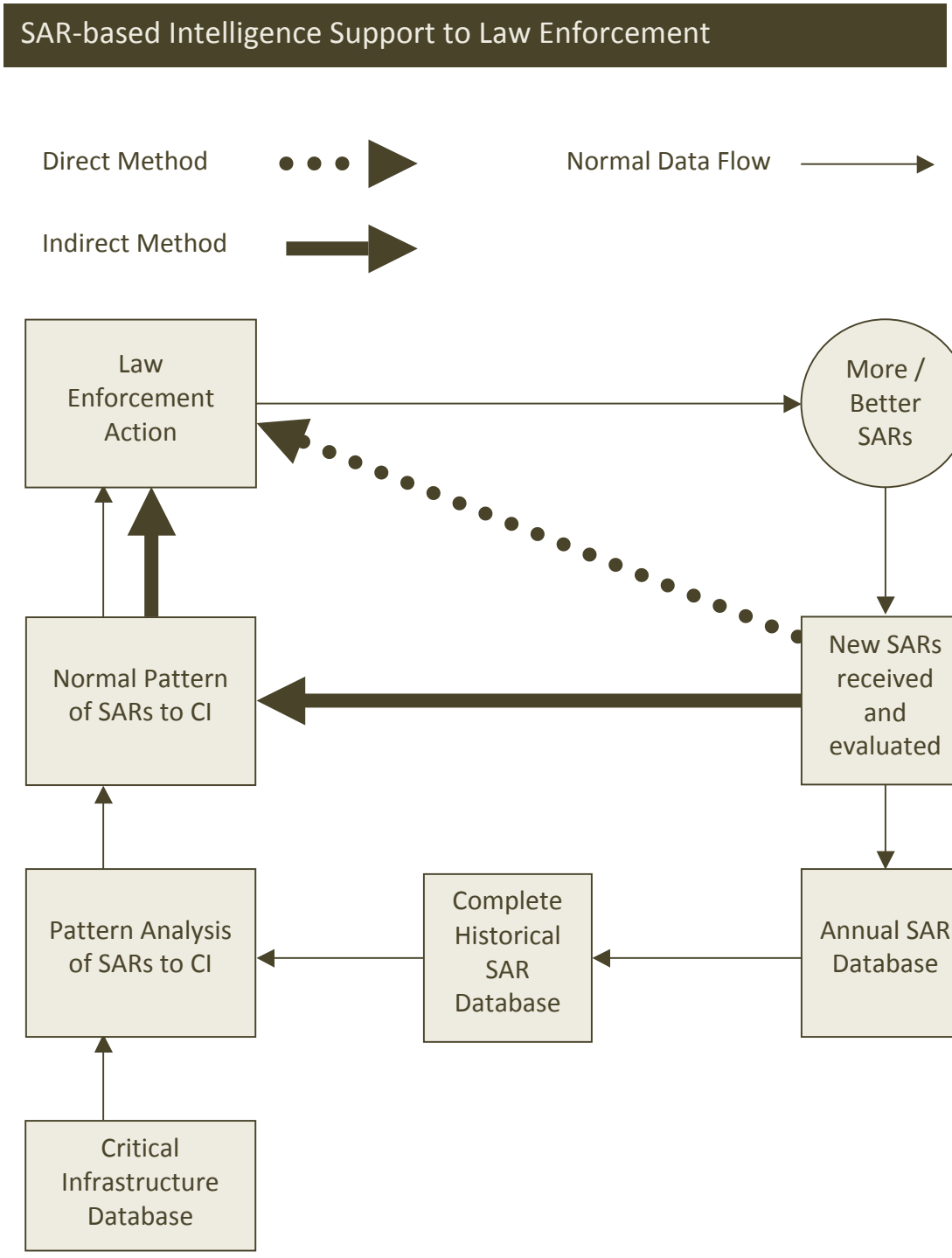

Figure 2. SAR-based Intelligence Support to Law Enforcement

Here is a brief summary of how such models and their results are generated. First, intelligence analysts format and evaluate all new SAR (a major task as noted above) and add them to the "complete historical SAR database." This database is then used with a similar (in that both are geo-rectified) database that holds information on critical infrastructure (CI) to feed the pattern analysis models, which, in turn, yield the "Normal Pattern of SAR" for each element of the CI. Critical Infrastructure analysts are responsible for generating the CI Database, which contains GIS and other information on each facility. Now for a closer look.

The Theoretical Foundation of the Indirect Approach

Given the obvious problems with using SAR as "intelligence," a number of professional intelligence officers have resisted spending scarce resources collecting and analyzing them. In the foreign intelligence world, a significant premium is placed on recruiting well-placed, reliable sources that can report raw intelligence that addresses specific collection requirements. Obviously, there is an effort by domestic law enforcement to utilize similar human sources (confidential informants) and technical collection (court approved wiretaps) to collect high-grade information. In the foreign arena, our intelligence agencies have been successful in collecting significant information on terrorist groups, their numbers, leaders, tactics, and techniques. This reporting has been used extensively in preparing analyses (target studies) at both the tactical and strategic levels. But within the U.S. domestic theatre, there just have not been many specific, credible reports on planned terrorist attacks within the U.S. since the 9/11 attacks. Although few in numbers, specific, credible intelligence reports have been critical in preventing terrorist attacks and both intelligence and law enforcement are actively seeking to acquire more such information.

On the other hand, there have been tens (possibly hundreds) of thousands of domestic suspicious activity reports collected in the U.S. by state and local law enforcement since 9/11. Statistically, the probability of any single SAR being an indicator of an actual terrorist plot is so small that it is insignificant. But the greater the number of SAR, the greater the overall probability that at least a few real indicators of threat exist within the total body of SAR reporting.

Most domestic intelligence practitioners characterize the counter-terrorism problem as separating the signal (the very rare true indicator of a terrorist threat) from the noise (the huge number of meaningless or benign SAR). For example, in 2009-2010 the FBI reports that of the 3,400 SAR entered into its system,[11] only fifty-six of these (less than 2 percent) merited an actual investigation.[12]

The indirect approach for using SAR to identify true threats harnesses the power of statistics and analysis to solve the signal/noise problem. Here is an example of the technique. Let's say there are 100 widget factories in the country and that we think (based on credible, specific intelligence regarding terrorist intent) that there is a near-term plot to attack one such factory. The challenge is to identify which particular widget factory is being targeted. Our understanding of how terrorists plan an attack tells us that operatives are likely to conduct at least a few surveillance operations before an attack. Further, we have collected a large number of SAR at or near the 100 widget factories over the past nine years – let's say 90,000 – roughly 100 SAR per plant per year. Obviously, if we are very lucky we will be able to identify the four or five of these 90,000

SAR that are true terrorist surveillance and immediately notify law enforcement and critical infrastructure protection to take appropriate actions.

But the likelihood of being so fortunate in identifying the actual SAR that are strong indicators of the targeted factory by examining each of the 90,000 relevant SAR is remote.

We have a much better chance of identifying a real indication of a threat if we characterize the challenge as defining the "normal" laydown of SAR at a widget factory with enough specificity to enable us to recognize a situation or SAR that is not normal. For example, if we can use our complete data set of SAR to develop a reliable expected pattern of SAR on any given target or area, then a significant deviation from that pattern is by definition an anomaly that deserves further investigation.

This approach is grounded in the "Law of Large Numbers."

In probability theory, the Law of Large Numbers (LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

For our widget factory, here is how the LLN might work in action. We take the full dataset of 90,000 SAR on widget factories and *if* this total number of SAR is large enough and *if* every facility is collecting and reporting SAR in a uniform and consistent fashion, then we should be able to develop a model which will tell us if there are an unusually large number or type of SAR at any given widget plant. We can do this because quantitative methodologists, working with intelligence and critical infrastructure analysts, can now construct a model that defines the normal pattern of SAR for each unique widget plant. This model actually forecasts the number and characteristics of SAR expected at each widget plant during a particular time period and a statistically significant deviation from this expected value would constitute an indicator of unusual (and possibly terrorist) activity. Obviously, from the LLN we know that the larger our complete database of SAR, the more reliable our "expected value." Note that intelligence analysts do not require any personal information on those conducting the suspicious activity to conduct their analysis.

While theoretically feasible, the practical problems involved in constructing and maintaining a SAR database in which there is consistency and uniform treatment of the collection and coding of the data (SAR) might well be impossible to overcome, at least in the near term. For example, the data set on SAR must be huge, relatively consistent, and comprehensive to enable robust analysis. One need only consider the fact that there are over 18,000 local, state, and federal law enforcement organizations within the U.S. collecting and reporting SAR to begin to understand and to size the obstacles to full and effective implementation. On the other hand, the cost of overcoming these difficulties and uncertainties must be counterbalanced against the potential payoff of preventing a successful terrorist attack, either by direct law enforcement action or through improved critical infrastructure (CI) protection.

SAR Support to Critical Infrastructure (CI) Protection

The second argument for a robust SAR program is its largely unrecognized role in CI protection. The DHS website states:

> CI (Critical Infrastructure/Key Resources) is an umbrella term referring to the assets of the United States essential to the nation's security, public health and safety, economic vitality, and way of life. CI is divided into 18 separate sectors, as diverse as agriculture and food, emergency services, and cyber networks. [See below.]
>
> Because this critical infrastructure provides our country with the enormous benefits, services and opportunities on which we rely, we (DHS) are very mindful of the risks posed to CI by terrorists, pandemic diseases and natural disasters. At the Department of Homeland Security, we know that these threats can have serious effects, such as cutting populations off from clean water, power, transportation, or emergency supplies.
>
> Secretary Napolitano is working to raise awareness about the importance of our nation's critical infrastructure, and strengthen our ability to *protect* it [emphasis added]. The Department oversees programs and resources that foster public-private partnerships, enhance protective programs, and build national resiliency to withstand natural disasters and terrorist threats.[13]

The 18 DHS Defined Critical Infrastructure Sectors	
Agriculture & Food	Emergency Services
Banking & Finance	Energy
Chemical	Government Facilities
Commercial Facilities	Information Technology
Commercial Nuclear Reactors, Materials, & Waste	National Monuments & Icons
Critical Manufacturing	Postal & Shipping
Dams	Telecommunications
Defense Industrial Base	Public Health & Healthcare
Drinking Water & Water Treatment Facilities	Transportation Systems

Figure 3. DHS Defined Critical Infrastructure Sectors

The National Infrastructure Protection Plan (NIPP) sets forth a comprehensive risk analysis and management framework and clearly defines critical infrastructure protection roles and responsibilities for DHS and other federal, state, local, tribal, and private sector CI partners. The NIPP provides the coordinated approach used to establish national priorities, goals, and requirements for infrastructure protection to ensure that funding and resources are applied effectively. The goal of the NIPP is to

build a safer, more secure, and more resilient America by enhancing protection of the nation's CI to prevent, deter, neutralize, or mitigate the effects of deliberate efforts by terrorists to destroy, incapacitate, or exploit them; and to strengthen national preparedness, timely response, and rapid recovery in the event of an attack, natural disaster, or other emergency.[14]

Just as in the intelligence support to law enforcement case, intelligence support to CI protection goes well beyond SAR-based information. National, homeland security, and law enforcement intelligence organizations all produce and disseminate reports on terrorist techniques, historical terrorist targets and tactics, and many other studies to help those responsible for protecting CI. In fact, such studies are used to develop and validate sets of indicators of pre-operational activity. But SAR-based intelligence also has an important role to play.

The central player in bringing intelligence to bear on CI protection at the national level is the DHS Homeland Infrastructure Threat and Risk Analysis Center (HITRAC) which conducts integrated threat and risk analyses for CI sectors. HITRAC is a joint fusion center that spans both the Office of Intelligence and Analysis (I&A) – a member of the Intelligence Community – and Infrastructure Protection. As called for in section 201 of the Homeland Security Act, HITRAC brings together intelligence and infrastructure specialists to ensure a sufficient understanding of the risks to the nation's CI from foreign and domestic threats.[15]

HITRAC, partnering with the National Infrastructure Simulation and Analysis Center (NISAC), has the mandate to lead in the development, testing, and evaluation of SAR-based models of "normal patterns" of SAR at CI. As noted above, HITRAC is unique in bringing together intelligence and critical infrastructure analysts at the national level. It also has been aggressive in working with state-level CI protection organizations and fusion centers to develop intelligence support to CI protection. On the other hand, NISAC has the sophisticated expertise needed to conduct statistical modeling. In fact, NISAC is a congressionally-mandated modeling, simulation, and analysis program.[16] The Center already prepares and shares analyses of CI including their interdependencies, consequences, and other complexities, so moving on to model the relationship between SAR and CI is a natural extension.

We can now move on to Figure 4, which displays the role of intelligence based on SAR in improving CI protection. Intelligence analysts are responsible for developing, updating, and improving the "complete historical SAR database." Similarly, critical infrastructure analysts are responsible for bringing their sector-specific expertise to bear in developing and maintaining a GIS-rectified CI database. These two teams of analysts, working in tandem, operate the models that yield a picture of the "normal pattern of SARs [*sic*] to specific CI facilities". This understanding of what constitutes a normal versus an abnormal situation at sector specific potential targets can then be used by the CI analysts to develop a more effective *protection* regime at that facility. Finally, in the process of advising CI stakeholders and partners, the CI analysts will encourage them to gather more and better reports of suspicious activity – a positive feedback loop strengthening the SAR process.

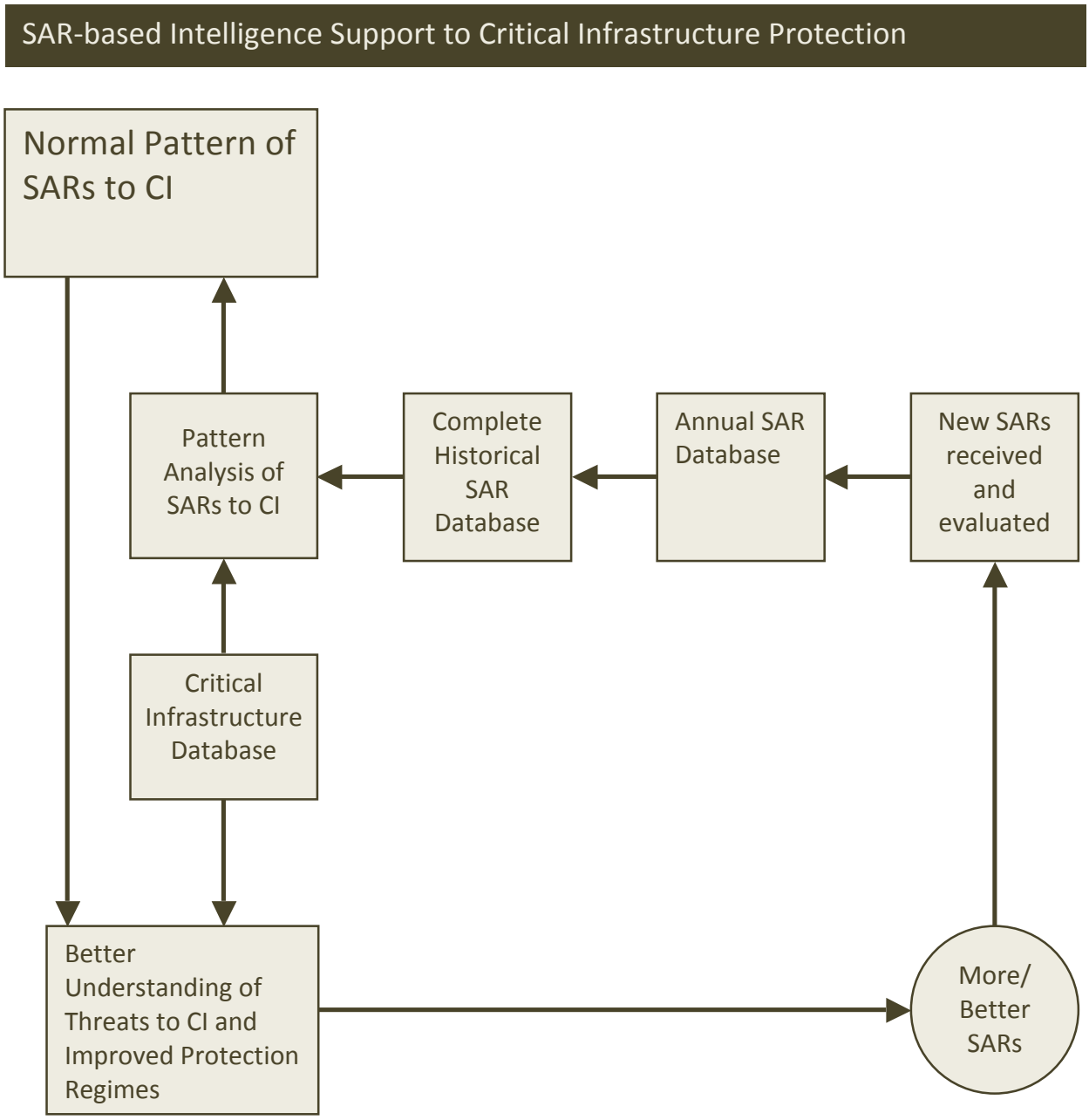

Figure 4. SAR-based Intelligence Support to Critical Infrastructure Protection

RECOMMENDATION

This paper has presented two SAR-based, data-intensive analytic techniques that have the *theoretic* potential to improve our counter-terrorism efforts by warning of terrorism activity and by assisting in the development of better CI protection regimes. Civil liberty activists point out the potential dangers to our privacy posed by a massive, national SAR program. They also question the effectiveness and efficiency of such a SAR effort.

Before national policymakers decide which way to go on this issue, it seems reasonable to conduct a validation test to determine whether or not the theoretic value-added of a robust SAR program can be proven in the real world. Such a validation program would begin with selection of a representative sample of the most important target sets (i.e. mass transit systems, bridges, dams, etc) and a concerted effort to collect a comprehensive data set of "suspicious activity" for this sample. Analysts from both the critical infrastructure and intelligence disciplines, supported by quantitative methodologists, would then be tasked to develop models of "normal" suspicious activity for each sample facility. If the models simply are not sensitive enough to detect simulated precursor terrorist activity or to provide insights into ways to improve protection of these facilities, then it is reasonable to constrain the SAR program. But if the models can do the job and actually are shown to constitute a unique, intelligence-driven capability to prevent terrorist attacks or at least to better protect our critical infrastructure, then further debate is in order before we simply discard that capability.

Dr. James E. Steiner is public service professor at Rockefeller College (SUNY Albany) where he teaches graduate courses in the craft of intelligence, with emphasis on intelligence analysis for homeland security. Dr. Steiner has more than forty years of experience conducting, leading, managing, teaching, and evaluating intelligence. After retiring in 2005 from a thirty-four year career at CIA, he taught intelligence analysis at the FBI Academy at Quantico. From 2006-2009 Dr. Steiner was an advisor both to the chief intelligence officer at the Department of Homeland Security and also to the director of New York State's Office of Homeland Security. He may be contacted at drjsteiner@gmail.com.

---

[1] U.S. Congress, House Committee on Homeland Security, Subcommittee on Intelligence, Information Sharing, and Terrorism Risk Assessment, *Homeland Security Intelligence: Its Relevance and Limitations*, Statement of Gregory T. Nojeim, Director on Freedom, Security, and Technology, Center for Democracy and Technology, 111th Cong., 1st sess., March 18, 2009, 1-2.

[2] Ibid., 14.

[3] Mark A. Randol, "Terrorism Information Sharing and the Nationwide Suspicious Activity Report Initiative: Background and Issues for Congress" CRS 7-5700 (Washington, DC: Congressional Research Service, November 5, 2009), 15-16, www.fas.org/sgp/crs/intel/R40901.pdf.

[4] Program Manager, Information Sharing Environment, "Findings and Recommendations of the SAR Support and Implementation Project" (Washington, DC: Bureau of Justice Assistance, US Department of Justice; the Major Cities Chiefs Association; DOJ's Global Justice Information, October 2008), 6.

[5] Ibid., 31.

[6] Program Manager, Information Sharing Environment, "Nationwide SAR Initiative Concept of Operations" (National SAR Initiative, December 2008), 8.

[7] "Final Report: Information Sharing Environment (ISE)—Suspicious Activity Reporting (SAR) Evaluation Environment" (Washington, DC: Bureau of Justice Assistance, U.S. Department of Justice, January 2010).

[8] Nojeim, testimony, *Homeland Security Intelligence*, 4-6.

[9] American Civil Liberties Union, "More About Suspicious Activity Reporting" (June 2010), http://www.aclu.org/spy-files/more-about-suspicious-activity-reporting

[10] Ibid.

[11] This is described at http://www.fbi.gov/page2/sept08/eguardian_091908.html.

[12] Joseph Straw, "Terror Threat Tracking System Shares Thousands of Tips from Locals, FBI Says" (March 2010), http://www.securitymanagement.com/print/6888.

[13] U.S. Department of Homeland Security, "Critical Infrastructure Protection" (August 2010), http://www.dhs.gov/files/programs/critical.shtm.

[14] U.S. Department of Homeland Security, *National Infrastructure Protection Plan: 2009*, http://www.dhs.gov/xlibrary/assets/NIPP_Plan.pdf

[15]U.S. Department of Homeland Security, "About the Homeland Infrastructure Threat and Risk Analysis Center (HITRAC)," http://www.dhs.gov/xabout/structure/gc_1257526699957.shtm

[16] The National Infrastructure Simulation and Analysis Center (NISAC), http://www.sandia.gov/nisac/