Dataveillance and the False-Positive Paradox

Javier Parra-Arnau, Claude Castelluccia

Dataveillance and the False-Positive Paradox

Javier Parra-Arnau, Claude Castelluccia

Dataveillance and the False-Positive Paradox

Javier Parra-Arnau, Claude Castelluccia

***Abstract*—In recent times, we are witnessing an increasing concern by governments and intelligence agencies to deploy mass-surveillance systems that help them fight terrorism. In this paper, we conduct a formal analysis of the overall cost of such surveillance systems. Our analysis starts with a fairly-known result in statistics, namely, the false-positive paradox. We propose a quantitative measure of the total cost of a monitoring program, and study a detection system that is designed to minimize it, subject to a constraint in the number of terrorists the agency wishes to capture. In the absence of real, accurate behavioral models, we perform our analysis on the basis of several simple but insightful examples. With these examples, we illustrate the different parameters involved in the design of the detection system, and provide some indicative and representative figures of the cost of the monitoring program.**

I. Introduction

The recent terrorist attacks have increased the level of online surveillance performed by governments around the world. In particular after the Charlie Hebdo attacks in Paris, the French government has voted a new signal intelligence law [1]. This new law includes the possibility to deploy black boxes in the network that systematically analyze the meta-data of Internet users in order to detect potential terrorists.

Although it is clearly legitimate for a government to actively fight terrorism, it is questionable whether the proposed large-scale surveillance solution is really rational and will increase security at all. In particular, it is questionable whether this solution is cost-efficient, since it is expected, according to the false positive paradox, that much of the resource will be spent analyzing the data of innocent people.

In this paper, we conduct a formal analysis of the overall cost of a surveillance system. Our analysis starts from the false-positive paradox, a statistical result where false positive events (i.e., identification of innocents as potential suspects) are more probable than true positive events (i.e., detection of terrorists). This paradox occurs when the number of events to be detected —terrorists in our case— is very small compared to the whole population. It results in very inefficient systems that produce highly unreliable results.

The aim is to understand under which conditions dataveillance may be rational and economical. Although mass-surveillance is also questionable in terms of privacy, we decided to avoid this debate and, instead, focus our analysis on the financial cost of such system. We formulate our analysis as an optimization problem: given a fixed limited budget, how we should allocate it in order to achieve the optimal results.

Javier Parra-Arnau and Claude Castelluccia and are with the Privatics research team, INRIA Grenoble - Rhône-Alpes (France),
E-mail: javier.parra,claude.castelluccia@inria.fr

Manuscript prepared May, 2015.

II. Background

In this section, we first describe our notation and terminology as well as the assumptions about the surveillance-system model. Secondly, we illustrate the false-positive paradox presented in the introductory section, which constitutes our starting point to analyze the cost of a mass surveillance.

A. Notation and Assumptions

Throughout this section, we shall follow the convention of using uppercase letters to denote random variables (r.v.'s), and lowercase letters to the particular values they take on. The measurable space in which an r.v. takes on values will be called an *alphabet*. With a mild loss of generality, we shall always assume that the alphabet is discrete. Probability mass functions (PMFs) are denoted by p, subindexed by the corresponding r.v. Accordingly, $p_X(x)$ denotes the value of the function p_X at x. We use the notations $p_{X|Y}$ and $p_{X|Y}(x|y)$ equivalently.

In the following, the r.v. X is used with full generality to include categorical or numerical *data* about an individual, although for mathematical simplicity we shall henceforth assume it models single-occurrence data, rather than tuples or sequences. X may represent, for example, the number of visited Web sites related to jihadism by an individual, the frequency with which this individual tweets fanatic, religious comments, or the number of social links that they actively maintain with extremists. Without loss of generality, we assume X takes on values on the alphabet $\{1,\ldots,n\}$.

In our analysis, we consider an ubiquitous surveillance system that gathers information about a population as a result of, for example, monitoring data and traffic on the Internet, tapping telephone lines and mining social-networking sites. Due to the tremendous amount of information involved, the surveillance system must necessarily *automate* the gathering of such data and their posterior analysis.

We assume a surveillance system that relies on automated computer software to carry out both tasks, and that reports to human investigators only when it detects patterns of individuals which can be classified as terrorists. Our analysis of the cost of this system assumes, however, that the cost of this automated, computerized monitoring is negligible compared to that of human inspection.

We believe this is a reasonable assumption, as it is consistent with some recent reports on the economic cost of such manual examination. As an example, the Canadian Security and Intelligence Service reportedly requires around 16 full-time personnel to install a listening device and about 33 agents to follow a suspect 24/7 for more than a week [2]. Also, in current practice, the number of human resources is what limits the ability of the surveillance system to fulfill its duties [3]. We have seen this in the past, for example in the case of the

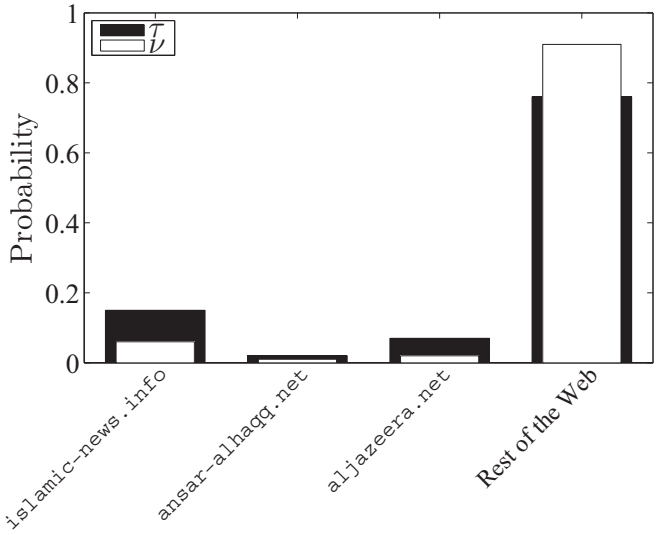


Fig. 1: We show a simple but illustrative example of the PMFs characterizing what might be regarded as terrorist and innocent Web-browsing patterns in a Muslim country. The profiles τ (terrorist) and ν (innocent) here depicted correspond to the probability distribution of the pages visited across the Web. Conceptually, these artificial profiles tell us that 15% of all page visits by a terrorist would be to the site `islamic-news.info`, whereas this would be just 6% in the case of an innocent.

Boston Marathon bombers and more recently with the Charlie Hebdo terrorists: they had been identified, but were not tracked anymore for budget and resource reasons.

That said, for the sake of illustration this work considers a simplified model for the automated detection process. More specifically, we assume that the monitoring system applies a binary hypothesis test [4, §11] to find out whether *one* observed data has been distributed according to a PMF τ that captures a *terrorist*'s characteristic behavior, or a distribution ν that reflects common patterns among *innocents*. We acknowledge, however, that a surveillance system will probably have more than one observation about an individual, and therefore will carry out the test on the basis of sequences of observed data, or equivalently, their corresponding empirical distributions.

We shall refer to those two distributions also as the terrorist and innocent *profiles*. Fig. 1 provides an example of such distributions that might reflect the Web-browsing habits of those profiles. We would like to stress that this example does not pretend to be an accurate or realistic representation of any actual profile. It merely aims to illustrate the kind of information captured by the distributions involved in the hypothesis-testing problem.

Let H be a binary r.v. representing the two possible hypothesis about the distribution of the observed data. Precisely, $H = 1$ with probability θ and $H = 2$ with probability $1 - \theta$, and X conditioned on H has PMF τ when $H = 1$ and ν when $H = 2$. A *randomized estimator* or *detector* $\hat{H}$ of H is a probabilistic decision rule determined by the conditional probability $p_{\hat{H}|X}$. The interpretation of such estimator is as follows: if X is observed to have value j, the detector concludes $H = 1$ with probability $p_{\hat{H}|X}(1|j)$, and $H = 2$ otherwise. Note that deterministic estimators are a particular case of randomized detectors.

The performance of a decision rule is typically characterized in terms of its error and detection probabilities. The probability of a false positive, which we shall also refer to as *misidentification rate*, is the probability that an innocent be considered as a terrorist by the surveillance system, that is, $p_{\hat{H}|H}(1|2)$. The probability of a true positive, on the other hand, is the probability of correctly identifying terrorists, i.e., $p_{\hat{H}|H}(1|1)$. We shall also regard this probability as the *accuracy rate*. The probabilities of false and true negatives are defined analogously.

B. The false-positive paradox

This section reviews a fairly known result in statistics, namely, the false-positive paradox [5].

Consider a very pessimistic or paranoic vision of the current situation of terrorism, in which a 0.1% of the population in France is terrorist. With a population of approximately 70 million people, we could regard this figure of 70 000 terrorists as a worst-case scenario or loose upper bound.

Suppose that the French intelligence agency Direction Générale de la Sécurité Intérieure has a surveillance system such as the one described in Sec. II-A, with access to a wide range of personal information including Internet, phone-call and banking records. In addition, assume that the agency has a highly accurate detection system at its disposal, with a true positive rate of 99% and a probability of false positive of 0.5%.

Bearing in mind all these assumptions, now we wonder about the reliability of a positive detection. In other words, given an individual who has been labeled as a terrorist by the automated detection system, what would be the probability that they have been correctly identified as such?

From Bayes' theorem, it is immediate to verify that such probability, which in information retrieval is called *precision* [6], yields

$$p_{H|\hat{H}}(1|1) = \frac{p_{\hat{H}|H}(1|1)\,\theta}{p_{\hat{H}|H}(1|1)\,\theta + p_{\hat{H}|H}(1|2)\,(1-\theta)} \simeq \frac{1}{6},$$

which means that, in average, just one out of six positive tests will in fact correspond to a terrorist. In absolute numbers, the number of false positives and true positives would be 349 650 and 69 300 respectively.

In the example given, the large number of false positives —compared to the number of true positives— appears to be contradictory: the accuracy and misidentification rates show a detection system that is certainly precise; but in reality, if an individual is classified as a terrorist, the probability that the system is correct is only $\frac{1}{6}$. This is known as the *false-positive paradox* and it arises when the overall population has a low incidence of terrorism (or of a health condition, in the case of medical testing) and this incidence rate is lower than the probability of a false positive.

The upshot is that, in a current state of affairs with fortunately very few terrorists —fewer in proportion to the false positive rate— there will more innocents misidentified as terrorists than terrorists correctly identified. In terms of resource efficiency, this has a straightforward consequence: the surveillance system will require to conduct an investigation on

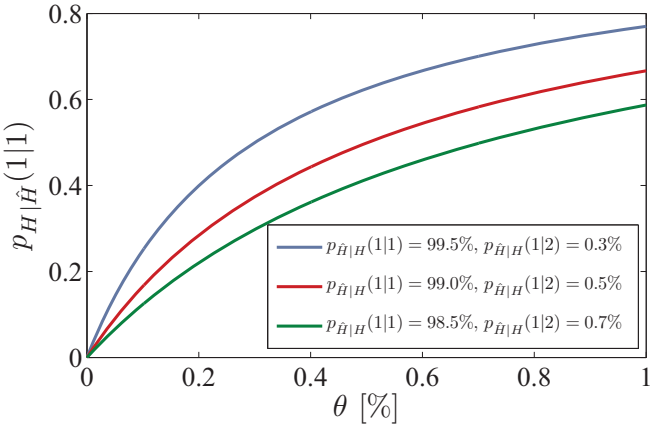


Fig. 2: The probability of a correct positive test is plotted for different values of accuracy and misidentification rates. Recall that the accuracy rate of a test is defined as the probability of correctly identifying terrorists, i.e., $p_{\hat{H}|H}(1|1)$. The rate of misidentification of a test is defined as the probability of incorrectly classifying innocents, that is, $p_{\hat{H}|H}(1|2)$. As noted from this figure, the false-positive paradox is observed when the ratio θ of terrorists to total population is lower than the misidentification rate.

the list of those people classified as terrorists when in most cases they will be innocent citizens.

Fig. 2 illustrates the impact that this result may have on a surveillance system, depending on the accuracy and misidentification rates of the test. In this figure, the probability $p_{H|\hat{H}}(1|1)$ is computed directly from Bayes' theorem, assuming no dependence among the different probabilities involved.

III. Detecting Terrorists

In the background section, we roughly evaluated the probability that an individual labeled as a terrorist be in fact a terrorist. For the sake of clarity, in that section we ignored the dependence between the accuracy and the misidentification rates. In this section, first we succinctly illustrate the variables that characterize these two rates; then we propose a quantitative measure of the economic cost of a mass surveillance program; and finally, we formulate a detector aimed at minimizing the overall cost of such monitoring. As we shall show in the coming sections, the purpose of designing cost-optimized detectors is to obtain, for several simple examples of τ and ν, some representative lower bounds on the expenditures of such surveillance programs.

Using basic probability theory, we can immediately check the relationship between the accuracy and the misidentification rates. For $i, j = 1, 2$, note that

$$p_{\hat{H}|H}(i|j) = \sum_{k=1}^{n} p_{\hat{H}|X}(i|k)\, p_{X|H}(k|j),$$

and in particular that

$$\begin{aligned} p_{\hat{H}|H}(1|1) &= \langle\, p_{\hat{H}|X}(1|\cdot), \tau \,\rangle, \\ p_{\hat{H}|H}(1|2) &= \langle\, p_{\hat{H}|X}(1|\cdot), \nu \,\rangle, \end{aligned} \tag{1}$$

where the symbol "$\langle\, , \,\rangle$" denotes the standard inner product on $\mathbb{R}^n$, and $p_{\hat{H}|X}(1|\cdot)$, τ and ν are interpreted here as vectors.

Two evident albeit insightful observations stem from (1). First, as fully expected, we note that the accuracy and misidentification rates depend on the reference, characteristic distributions that describe the behaviors of a terrorist and an innocent. Secondly, we observe that both rates are subject to the decision rule that concludes when an individual is a terrorist. The reference distributions are the result of profiling both terrorists and innocents based on the available data, and can be seen as input parameters of the estimator. The decision rule, on the other hand, is determined by the particular optimality criterion chosen to design it.

Some classical optimality criteria are the Bayes, minimax and Neyman-Pearson designs [7]. Their formulations of the hypothesis test between $H = 1$ and $H = 2$ are given, respectively, by

(i) $\min_{p_{\hat{H}|X}} \theta\, p_{\hat{H}|H}(2|1) + (1-\theta)\, p_{\hat{H}|H}(1|2)$,
(ii) $\min_{p_{\hat{H}|X}} \max\{p_{\hat{H}|H}(2|1), p_{\hat{H}|H}(1|2)\}$,
(iii) $\min_{p_{\hat{H}|X}} p_{\hat{H}|H}(2|1)$ subject to $p_{\hat{H}|H}(1|2) \leqslant \gamma$.

In this work, we contemplate an optimality criteria for the automated surveillance system's detector, which might resemble, to some extent, the Bayes and Neyman-Pearson formulations. In particular, we consider minimizing a measure of the overall economic cost of such system for a desired probability of true positives, and vice versa. The next two subsections elaborate on this measure of economic cost and provide the formulation of the proposed detector.

A. Measure of the Overall Economic Cost

Our measure of the total cost of a surveillance system revolves around the false-positive paradox and the fact that each alarm requires a costly investigation to decide whether it is real or not.

As mentioned in Sec. II-B, one of the consequences of the reduced incidence of terrorists is that the monitoring system will require to conduct a further investigation on the set of individuals labeled as terrorists, a set with unluckily a large number of false positives compared to the number of true positives. We denote this set of "still-suspects" by S.

We assume that the agents of the intelligence service who are in charge of examining this set proceed in a serial manner, that is, they do not undertake a new investigation until an individual is correctly identified as innocent or terrorist — this can be justified in terms of the limited although often disproportionate number of resources available to an agency. Our analysis of the economic cost per suspect, however, is valid regardless of the number of agents working in parallel on a partition of S, under the reasonable assumption that simple random sampling is employed.

Given a subset of S, the number of individuals that an agent will have to examine until they catch a terrorist is clearly a geometric r.v. with parameter $p_{H|\hat{H}(1|1)}$. We denote by $\mathcal{C}_c$ the *cost of checking* the true condition of one particular individual within S. This would include, for example, the costs of tracking, detaining and interrogating them[1]. Recall that the expected value of a discrete r.v. geometrically distributed is

[1] As a reference, [8] provides a rigorous estimate of the costs of different location-tracking techniques.

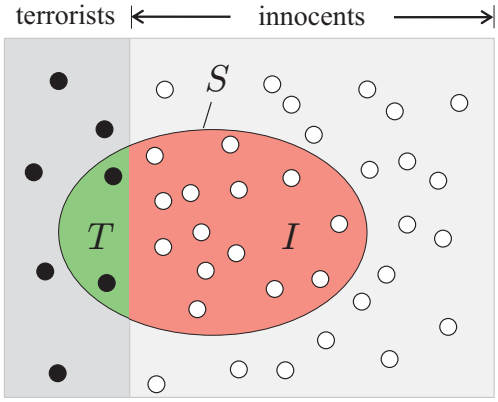


Fig. 3: The set of suspects or positive tests S is composed of those individuals who have been classified as terrorists by the detector. We propose measuring the total cost of a surveillance system essentially as the cardinality of S. The cost per terrorist, on the other hand, is roughly defined as the ratio of positive tests to true positives, i.e., $|I \cup T| \,/\, |T|$.

the inverse of its parameter. Accordingly, for $\theta > 0$ we define the *cost per terrorist* $\mathcal{C}_t$ of a security agency as

$$\mathcal{C}_t = \frac{\mathcal{C}_c}{p_{H|\hat{H}}(1|1)}.$$

The *total cost* $\mathcal{T}$ of the surveillance program is defined intuitively by multiplying $\mathcal{C}_t$ by the number of true positives. We would like to emphasize that this latter measure of cost is also independent of the number of agents involved in the inspection of S. The number of human resources to this end has only an impact on the time taken by this examination.

The proposed measures of cost can also be interpreted in terms of the sets of false positives and true positives. This is illustrated in Fig. 3, where we denote those sets by I and T, respectively. Consistently with the false-positive paradox, this figure shows a set I much larger than T. Recall that we defined S as the set of suspects, that is, $S = I \cup T$. Accordingly, it is straightforward to check that $\mathcal{C}_t/\mathcal{C}_c = |S| \,/\, |T|$ and that $\mathcal{T}/\mathcal{C}_c = |S|$. In a nutshell, we may regard our measure of total cost essentially as the number of suspects listed by the automated detector.

B. Cost-Optimized Detection

Having defined a metric of the overall economic cost of a surveillance system, a detector may be designed accordingly to minimize it, maybe accompanied with some constraints on the probabilities of error and detection. In this work, we contemplate two equivalent design principles for such detector. Our first design considers the case in which the intelligence agency wishes to minimize the total cost of its surveillance system, while ensuring that a target, minimum percentage of terrorists $\kappa_{\min}$ is captured. The formulation of the corresponding detector is given by

$$\min_{p_{\hat{H}|X}} \theta\, p_{\hat{H}|H}(1|1) + (1-\theta)\, p_{\hat{H}|H}(1|2) \qquad \text{subject to } p_{\hat{H}|H}(1|1) \geqslant \kappa_{\min}. \quad (2)$$

We shall denote by $\mathcal{T}^*$ the *minimum total cost* attained by said detector. We hasten to stress that, in the trivial case when $\theta = 0$, a positive $\kappa_{\min}$ does not make sense. On the contrary, when $\theta > 0$, a target percentage of terrorists to be caught of $\kappa_{\min} = 0$ leads clearly to a minimum total cost $\mathcal{T}^* = 0$.

TABLE I: Description of the variables used in our analysis.

Symbol	Description
X	*observed data* about an individual
n	*cardinality* of the alphabet of X
$H, \hat{H}$	*true condition* (i.e., terrorist or innocent) of an individual, and *estimated condition* by the surveillance system's detector
τ, ν	terrorist and innocent *reference profiles*
θ	ratio of terrorist to total population
$\mathcal{C}_c$	the *cost of checking* is defined as the cost of carrying out a human investigation to ascertain the true condition of a suspect
$\mathcal{C}_t$	the *cost per terrorist* is defined as the product of $\mathcal{C}_c$ and the average number of suspects that an agent must examine to capture one terrorist
$\mathcal{T}$	the *total cost* is defined as the product of $\mathcal{C}_t$ and the number of true positives
$\kappa_{\min}$	*minimum* percentage of terrorists the surveillance system aims at *capturing*
$\beta_{\max}$	*maximum budget* of the surveillance system to conduct human investigations on the list of suspects
I, T, S	sets of false positives, true positives and positive tests

Our second design approaches the complementary case to (2). That is, it considers a scenario in which the security agency wants to maximize the number of captured terrorists for a maximum budget $\beta_{\max}$. Said otherwise, we simply contemplate exchanging the objective and the constraint functions of the previous detector.

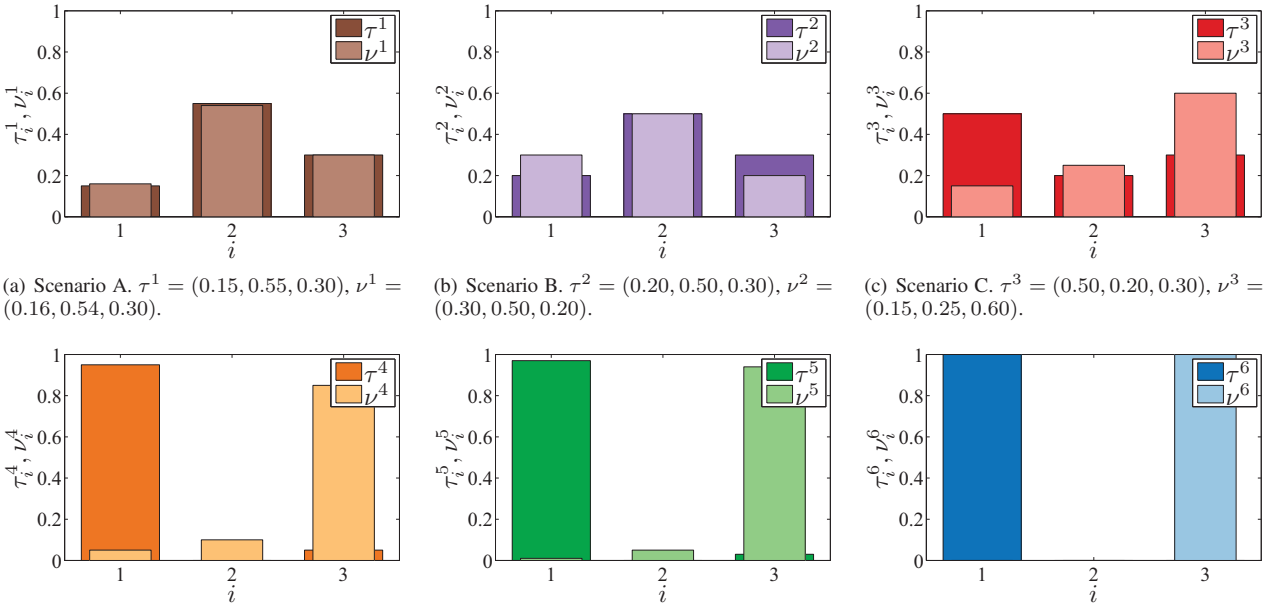
In this case, the formulation of the associated detector is given by the optimization problem

$$\max_{p_{\hat{H}|X}} p_{\hat{H}|H}(1|1) \qquad \text{subject to } \mathcal{T} \leqslant \beta_{\max}.$$

It can be shown, however, that this latter optimization problem characterizes the *same* optimal trade-off between total cost and accuracy rate described by (2). Consequently, because the analysis of either of these two formulations will yield entirely equivalent results, next subsections will just refer to (2) for simplicity. Table I provides a summary of the notation used in this section.

IV. Numerical Examples

This section illustrates the different parameters involved in the design of a detection system, and provides some indicative and representative figures of the cost of a mass monitoring program. In the absence of real, accurate data that allow us to characterize mainly the terrorist's profile —but also the innocent's—, we conduct our analysis on the basis of some simple but insightful examples of such distributions. We would like to emphasize that, accordingly, the results shown here are not a precise, detailed characterization of the actual cost of a real surveillance system. As we shall see in the next subsections, this is strongly influenced by the behavioral models assumed.



(a) Scenario A. $\tau^1 = (0.15, 0.55, 0.30)$, $\nu^1 = (0.16, 0.54, 0.30)$.

(b) Scenario B. $\tau^2 = (0.20, 0.50, 0.30)$, $\nu^2 = (0.30, 0.50, 0.20)$.

(c) Scenario C. $\tau^3 = (0.50, 0.20, 0.30)$, $\nu^3 = (0.15, 0.25, 0.60)$.

(d) Scenario D. $\tau^4 = (0.95, 0, 0.05)$, $\nu^4 = (0.05, 0.10, 0.85)$.

(e) Scenario E. $\tau^5 = (0.97, 0, 0.03)$, $\nu^5 = (0.01, 0.05, 0.94)$.

(f) Scenario F. $\tau^6 = (1, 0, 0)$, $\nu^6 = (0, 0, 1)$.

Fig. 4: Examples of probability distributions for terrorists τ and innocents ν, sorted from (a) to (f) in decreasing order of similarity.

The section is organized as follows. First, we introduce in Sec. IV-A two common metrics of distribution distance. Then, Sec. IV-B shows the reference profiles that will be used throughout our analysis. And finally we present our results in Sec.IV-C.

A. Measures of Profile Similarity

As we shown in Sec. III, the performance of a detector is determined by the reference distributions τ and ν that model terrorist and innocent behaviors. Although not explicitly stated in that section, it is clear that the error and detection probabilities, as well as the corresponding cost of the surveillance system, will largely depend on to what extent these two profiles diverge. Intuitively, the more dissimilar these profiles are, the less is the probability of incorrectly identifying individuals and hence the cost of the system. To quantify the impact of profile similarity on said cost, this section briefly introduces two measures of statistical distance between distributions, namely the cosine distance and the Kullback-Leibler (KL) divergence.

The *cosine distance* [9] is a simple and robust measure of dissimilarity between vectors, and is defined as

$$\mathrm{C}(\tau \,\|\, \nu) = 1 - \frac{\langle \tau, \nu \rangle}{\|\tau\| \|\nu\|},$$

where $\|\cdot\|$ denotes the Euclidean norm. It is important to notice that the cosine distance is not a proper metric as it does not satisfy the triangle inequality. Nevertheless, it does provide a measure of distance: in the case of probability distributions, it ranges from 0, meaning the PMFs are identical, to 1, indicating orthogonality.

The second distance measure we shall utilize to explore the effect of profile similarity on the cost of a surveillance system is the *KL divergence*, a fundamental quantity in information theory that arises, for instance, in the (optimal) likelihood ratio test of the Neyman-Pearson formulation [4]. The KL divergence between τ and ν is defined as

$$\mathrm{D}(\tau \,\|\, \nu) = \sum_{i=1}^{n} \tau_i \log \frac{\tau_i}{\nu_i}.$$

When not specified, the base of the logarithms is taken to base 2. Exactly as with the cosine distance, the KL divergence is not a metric as it is neither symmetric nor satisfies the triangle inequality. It provides, however, a measure of discrepancy between distributions, in the sense that $\mathrm{D}(\tau \,\|\, \nu) \geqslant 0$, with equality if, and only if, $\tau = \nu$.

In the definition of divergence, we shall use the convention that $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{\nu} = 0$ and $\tau \log \frac{\tau}{0} = \infty$. This can be justified by continuity arguments. As we shall show in Sec. IV-C, this information-theoretic quantity will allow us to capture the special case when $\tau_i > 0$ and $\nu_i = 0$ for some $i = 1, \ldots, n$, for which $\mathrm{D}(\tau \,\|\, \nu) = \infty$.

B. Scenarios

Our analysis is carried out for some examples of reference profiles, under the assumption that these profiles are modeled as PMFs, that is, as histograms of relative frequencies of data (from terrorists and innocents) within an arbitrary alphabet. A fairly realistic example of reference profiles could be the percentage of time spent by terrorists and innocents in some specific Web pages related to jihadism.

Throughout this series of examples, we shall assume a surveillance system monitoring a population of 70 million people, like in France approximately. The results presented in next section are shown for the six pairs of distributions $(\tau^j, \nu^j)_{j=1}^6$ depicted in Fig. 4. The number of data categories

	Scenarios					
Statistical distance	A	B	C	D	E	F
Cosine distance	0.0002	0.0263	0.2583	0.8896	0.9585	1.0000
KL divergence	0.0006	0.0585	0.5041	3.8312	6.2528	∞

Fig. 5: We show the approximate values of *dissimilarity* between the terrorist and innocent distributions plotted in Fig. 4.

considered in our analysis is $n = 3$. In terms of the example of profiles mentioned above, $i = 1, 2$ could be two pages related to Islamic extremism, and $i = 3$ the rest of the Web.

The pairs of profiles represented in that figure merely attempt to reflect some plausible scenarios in regards to the similarity between the two PMFs. Although in Sec. IV-A we introduced two quantitative measure of discrepancy between distributions, we shall also refer to those scenarios in qualitative terms, from low (scenario A) to high (scenario F) dissimilarity, consistently with the fact that $\mathrm{C}(\tau^i \,\|\, \nu^i)$ and $\mathrm{D}(\tau^i \,\|\, \nu^i)$ are strictly increasing with i. Table 5 shows the values of dissimilarity between each of the pairs of profiles assumed in our analysis.

In all the examples shown, the optimization problem inherent in the definition of the detector (2) has been computed numerically. The numerical method chosen is the sequential quadratic programming optimization algorithm [10], implemented by the Matlab R2014a function `fmincon`.

C. Results

We shall start our analysis by examining the example of distributions shown in Fig. 4(c), which represents an intermediate case in terms of profile similarity—a cosine distance of approximately 0.2583 seems to indicate this. We consider a ratio of terrorists to total population $\theta = 0.1‰$, and assume that the intelligence agency wishes to capture at least $\kappa_{\min} = 75\%$ of them. Interpreted as an $\mathbb{R}^{2\times 3}$ matrix, the detector that minimizes the total cost for the pair (τ^3, ν^3) is given approximately by

		X		
		1	**2**	**3**
$\hat{H}$	**1**	1.000	1.000	0.167
	2	0.000	0.000	0.833

where the entry i, j is the probability of deciding $\hat{H} = i$ when $X = j$ is observed. The estimator above has been computed numerically, using the method mentioned in Sec. IV-B, and tells us that an individual will be considered innocent only when $X = 3$, and with probability 0.833. The corresponding accuracy and misidentification rates can be obtained straightforwardly from this matrix and the reference distributions, through the expression (1). They yield 0.75 and 0.50, respectively. Recall that $\mathcal{T}/\mathcal{C}_c$ is the number of suspects that the agents of the surveillance system will have to further investigate on an individual basis. For the estimator at hand, this number reaches the high figure of 35 001 750 individuals, just about half of the population assumed.

The low-similarity cases E and F will show next that this enormous cost, given above in number of suspects, is in part a consequence of the semblance between τ^3 and ν^3. Fig. 4(e)

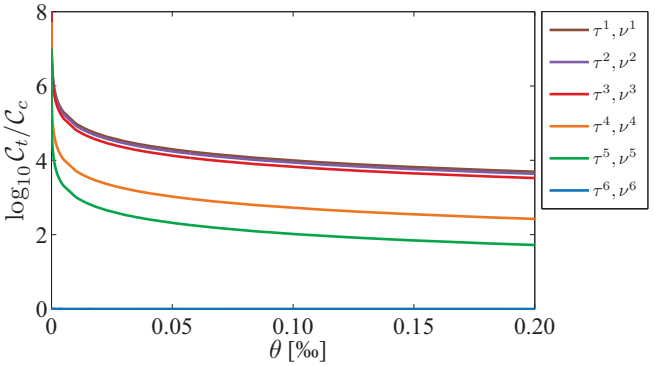


Fig. 6: We plot the minimum cost per terrorist, given in number of suspects per true positives, as a function of the ratio θ of terrorists to total number of individuals. We assume that the intelligence agency aims at capturing, at least, $\kappa_{\min} = 75\%$ of those terrorists.

represents two distributions such that $\nu_2 > \tau_2 = 0$. Intuitively, this means that, if $X = 2$, the individual in question will never be regarded as a terrorist, which will have an impact on the misidentification rate and therefore on the total cost of the system. For the same values of θ and $\kappa_{\min}$, and for the PMFs (τ^5, ν^5), the estimator that solves (2) is given by the matrix

		X		
		1	**2**	**3**
$\hat{H}$	**1**	0.773	0.000	0.000
	2	0.227	1.000	1.000

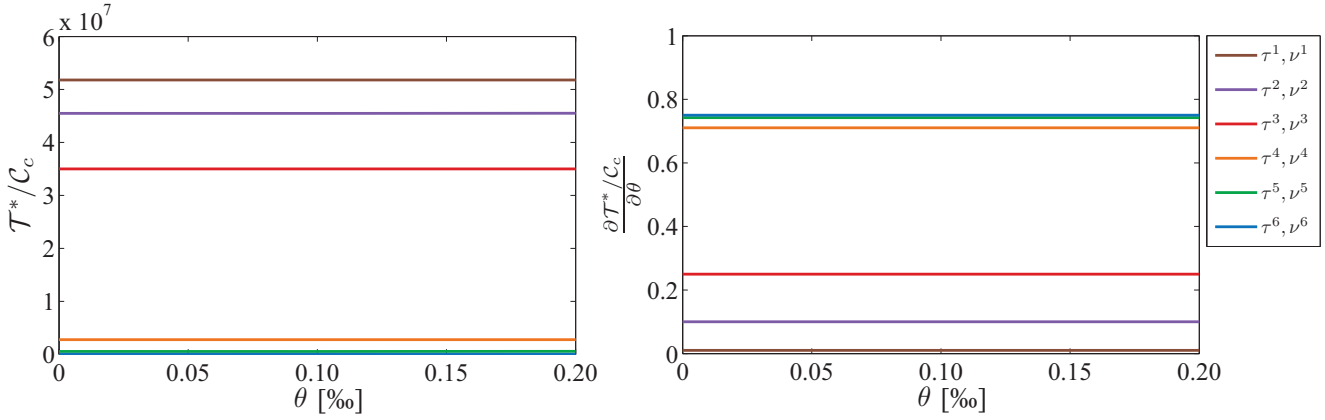
An interesting observation that follows from this detector is the relatively low value of $p_{\hat{H}|X}(1|1)$. At first sight, this may seem surprising since the probability of observing $X = 1$ for a terrorist (97%) is much higher than that for an innocent (1%). However, this is an expected consequence of the optimality criterion chosen. To illustrate this, simply note that a probability $p_{\hat{H}|X}(1|1) > 0.773$ would imply, from (1), an accuracy rate $p_{\hat{H}|H}(1|1) = p_{\hat{H}|X}(1|1)\tau_1^5 > \kappa_{\min}$. Such a detector would obviously be feasible, on account of the inequality constraint imposed in (2). But certainly, it would give a greater number of true positives and thus would incur a higher cost.

As anticipated, the performance of this detector is much better than that of scenario C. The misidentification rate is reduced significantly, yielding $p_{\hat{H}|X}(1|1)\,\nu_1^5 = 0.773\%$, and the minimum number of suspects to examine $\mathcal{T}^*/\mathcal{C}_c$ drops to 546 433 individuals. This latter figure represents 1.56% of the total cost of case C.

Not entirely unexpectedly, these results are further enhanced by the extreme case F, which we represent in Fig. 4(f). For the same $\theta = 0.1‰$ but for a more stringent requirement $\kappa_{\min} = 1$, the optimal estimator is given by the matrix

		X		
		1	**2**	**3**
$\hat{H}$	**1**	1.000	0.000	0.000
	2	0.000	0.000	1.000

which concludes that an individual is a terrorist if, and only if, $X = 1$. Naturally, $\mathcal{T}^*/\mathcal{C}_c$ attains the total number of terrorists, 7 000, and the probabilities of true-positive and false-positive tests become 1 and 0 respectively. Simply put, the surveillance system does not misidentify any terrorist and any innocent.



(a) (b)

Fig. 7: Minimum total cost, given in number of suspects, and its derivative with respect to the ratio of terrorists (per mille) for $\kappa_{\min} = 75\%$.

Compared to case E, this implies a reduction of 98.72% in total cost. The relative increment in cosine distance between the scenarios E and F, however, is only 4.33%, which seems to indicate that this said distance is not suitable to quantify significant differences in overall cost. The KL divergence, on the contrary, appears to capture this notable reduction in $\mathcal{T}^*$.

Recall that $\mathcal{C}_t/\mathcal{C}_c$ is the ratio of positive tests (i.e., individuals labeled as terrorists by the automated detection system) to true positives tests, that is, the inverse of precision. For a fixed $\kappa_{\min} = 0.75$, Fig. 6 shows the value of such ratio, which results from dividing the minimum total cost $\mathcal{T}^*/\mathcal{C}_c$ by the number of true positives. Later on we shall examine this minimum total cost for a range of values of $\kappa_{\min}$. The displayed results correspond to the six pairs of distributions represented in Fig. 4.

Consistently with our previous observations about the total-cost metric, we note that those cases in which the differences between τ and ν are more pronounced lead to lower minimum costs per terrorist. In all but case F, we also observe a ratio $\mathcal{C}_t/\mathcal{C}_c$ that increases exponentially as θ approaches 0; clearly, our measure of cost per terrorist is not defined at this extreme value. The behavior of these ratios are in line with the conclusions drawn in Sec. II-B about the large number of false positives—compared to the number of true positives. As an example, for $\theta \leqslant 0.1‰$ we notice that cases A-E would require examining, in average, a minimum of 100 suspects to capture one terrorist. The extreme case F, then again, yields an expected minimum ratio of 1 positive test per terrorist, regardless of the ratio of terrorists assumed.

Fig. 7, on the other hand, shows the dependence of the minimum total cost $\mathcal{T}^*/\mathcal{C}_c$ on the ratio of terrorists. The curves are also plotted for $\kappa_{\min} = 75\%$ and $\theta \in (0, 0.0002]$. Although $\mathcal{T}$ is defined for $\theta = 0$, this figure does not represent such values since we are considering a positive $\kappa_{\min}$.

Fig. 7(a) confirms the intuitive, preliminary findings suggested at the beginning of this section: the higher the similarity between the reference distributions, the higher the minimum overall cost. The results indicate that $\mathcal{T}^*$ is roughly linear with the ratio of terrorists. Fig. 7(b) plots the derivative of $\mathcal{T}^*/\mathcal{C}_c$ with respect to θ, which, in addition, allows us to appreciate the relatively low rates of increase guessed in Fig. 7(a). An interesting observation arising from this former figure is that an increase in the ratio of terrorists $\Delta\theta = \delta$ leads to an increase in the minimum total cost $\Delta\mathcal{T}^* < \delta$. Informally, this means that a possible overestimate of the ratio of terrorists by the intelligence agency would not have a great impact in terms of cost—at least for the distributions considered here. Lastly, consistently with Fig. 6, we check that case F yields $\mathcal{T}^*/\mathcal{C}_c = 5\,250$, that is, the total number of terrorists the agency aims at detaining.

The optimal trade-off between total cost and accuracy rate is illustrated in Fig. 8 for $\theta = 0.1‰$. The figure plots, more specifically, the minimum quotient $\mathcal{T}/\mathcal{C}_c$ as a function of the minimum percentage of terrorists $\kappa_{\min}$ the surveillance system wants to capture. Conversely, for a maximum budget $\beta_{\max}/\mathcal{C}_c$ (in number of suspects to interrogate), it gives us the maximum number of terrorists detained.

The trade-off curves are plotted again for the six examples of reference profiles. Remarkably enough, we observe that such curves exhibit a convex, piecewise-linear behavior, analogously to the receiver operating curve of the Neyman-Pearson test [10]. This observation has an evident practical implication. In the scenario C, for instance, we see that an increase of $1‰$ in $\kappa_{\min}$ (i.e., 7 more terrorists) leads to examining 212 more suspects; this is for $\kappa_{\min} \in [0, 0.50]$. However, this same increase in the interval $[0.69, 1]$ translates into 1 398 new suspects.

On the other hand, Fig. 8 shows relatively low values of total cost in those cases where the corresponding profiles are more dissimilar. For example, case E, with $\mathrm{D}(\tau^5 \,\|\, \nu^5) \simeq 6.2528$, yields only $\mathcal{T}^*/\mathcal{C}_c = 36\,310$ individuals for $\kappa_{\min} = 0.50$. The scenario with the most similar profiles, in contrast, indicates that half of the population should be investigated. Also, we note that the quotient $\mathcal{T}^*/\mathcal{C}_c$ in cases A-C becomes the size of the whole population when $\kappa_{\min} = 1$. In cases D and E such quotient is slightly reduced, being approximately 90% and 95% of the total number of individuals. Finally, in accordance with our previous observations about the cost per terrorist, case F gives the total number of terrorists.

Our last figure, Fig. 9, illustrates the impact that profile similarity, quantified through the two distance measures introduced in Sec. IV-A, may have on total cost. In particular, this figure

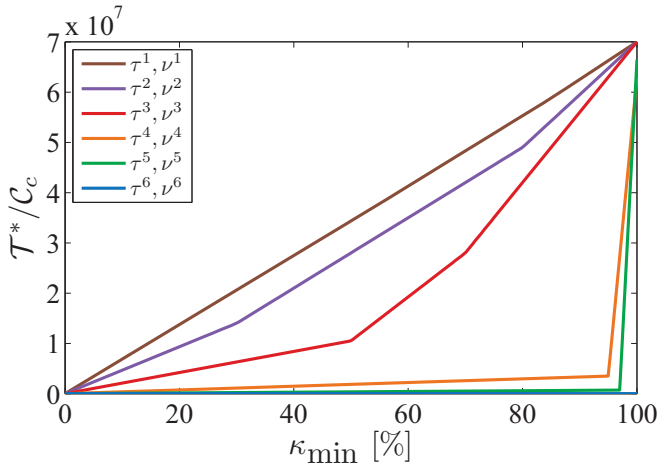


Fig. 8: Optimal trade-off between the cost of a surveillance system, given in number of individuals to investigate, and the minimum percentage of terrorists this system is required to capture. In this figure, we assume a ratio of terrorists to total population $\theta = 0.1‰$.

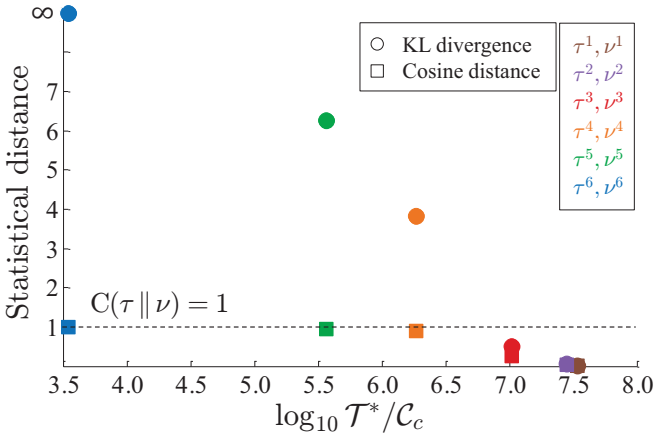


Fig. 9: We plot profile dissimilarity versus the minimum quotient $\mathcal{T}/\mathcal{C}_c$ for $\kappa_{\min} = 0.50\%$, $\theta = 0.1‰$ and for the six reference distributions considered in this section. Profile dissimilarity is measured with the KL divergence and the cosine distance.

shows the minimum $\mathcal{T}/\mathcal{C}_c$ versus the KL divergence and the cosine distance, for $\theta = 0.1‰$ and for each of the examples of distributions plotted in Fig. 4. We assume $\kappa_{\min} = 0.50\%$ and, consequently, the abscissa ranges from approximately half of the number of terrorists to about half of the overall population.

Apart from the fact that the minimum total cost decreases with these two measures of profile dissimilarity, an immediate observation is that the KL divergence is more sensitive to cost differences than the cosine distance. For example, from case E to D, $\mathcal{T}^*$ increases approximately 406.58%, whereas the cosine distance experiences a relative reduction of just 7.19%. This is in contrast to KL divergence, which is dropped 38.73%.

Another interesting remark is that this latter measure may capture the interesting scenario when $\tau_i > 0$ and $\nu_i = 0$ for some $i = 1, \ldots, n$, for which $\mathrm{D}(\tau \,\|\, \nu) = \infty$. Precisely, this corresponds to case F, which gives the minimum attainable total cost, 3 500. We must hasten to stress, however, that this does not imply that any pair of distributions satisfying the above requirements will lead to this minimum total cost value. For example, the KL divergence between the distributions $\tau = (0, 0.3, 0.7)$ and $\nu = (0.3, 0, 0.7)$ is infinite, but $\mathcal{T}^*/\mathcal{C}_c$ is around 14 million of positive tests.

The cosine distance, on the other hand, does reflect, in a more general way, this sense of orthogonality observed in the scenario F, which confirms the intuition that orthogonal profiles may lead to minimum attainable total costs and hence calls for the exploration of behavioral models satisfying this property.

V. Conclusions

This paper aims to provide insight into the economic cost of a surveillance system. Our analysis starts with the false-positive paradox, a fairly-known result in statistics in which false positives are more likely than true positives.

Our work does not evaluate the actual cost of a monitoring program —since the real distributions are not available in general—, but provides some indicative and representative figures of its cost on the basis of some simple but insightful examples of such distributions. The analysis is also conducted under the assumption that the surveillance system estimates the condition of an individual based on a single observation, rather than sequences of observed data. The simplicity of the model assumed allows us to show the relationships among the design parameters of such system in an illustrative manner. The main contributions of this work are, more specifically, a quantitative measure of the overall economic cost of a monitoring system, and the illustration of the optimal trade-off between this measure of cost and the requirements of this system in terms of the percentage of terrorists it wishes to capture.

The main results of this analysis are summarized next:

- We show that the proposed metrics of total cost and cost per terrorist are strongly dependent (or sensitive) on the feature distributions of terrorists and innocents.
- The optimal trade-off between overall cost and accuracy rate is observed to be convex, piecewise linear, which resembles the receiver operating curve of the Neyman-Pearson test.
- We observe an extremely large number of false positives, except in the scenario where the distributions of terrorists and innocents are orthogonal, as effectively captured by the cosine distance. In order to be orthogonal, the terrorist profiles and features have to be unique and very distinctive from other people profiles. Defining such profiles is challenging since scientists do not have access to the data of many terrorists. Besides, current results tend to show that terrorists have personality traits that are indistinguishable from traits of the general population [11]. Also, it is very likely that terrorists will use tools such as encryption tools or proxies, in order to perturb their profiles.
- Our results show that the total cost increases linearly with the ratio of terrorists, but the rate of increase is relatively low in the six scenarios considered. As depicted in Fig. 7, the total cost is similar regardless of the percentage of terrorists. This is a quite interesting observation because this means that, when the security agency has to decide the

budget, it will not need to be very accurate in estimating the percentage of terrorists within the population. On the other hand, this figure also shows that the efficiency of the system increases with the number of suspects, but is very low when the number of terrorist is small compared to the population size, which is fortunately the case. Mass surveillance of the entire population is logically sensible only if the number of persons to identified is high, which happens in McCarthy-type national paranoia or political espionage [12].

In closing, this paper demonstrates that dataveillance is not a very economical solution to fight against terrorism. More false positive will only overstress technologies, thus causing even more work for signals-intelligence agents, who are already overloaded [13]. In fact, the Charlie Hebdo terrorists were known by the French security agency prior to their attack. They were not followed and tracked anymore for budget and resource reasons. One might wonder how a dataveillance system that generates so many false positive, and is so easy to circumvent, will help improving the situation.

Lastly, as a future research line we plan to extend our cost analysis to the case when the automated detection process relies on sequences of observed data. In this same direction, we intend to analyze the few terrorist databases publicly available with the aim of building approximate behavioral models that better reflect the cost of a mass surveillance system.

REFERENCES

[1] A. J. Rubin and D. E. Sanger, "Familiar swing to security over privacy after attacks in france," May 2015, accessed on 2015-05-09. [Online]. Available: http://www.nytimes.com/2015/05/07/world/europe/france-expanded-surveillance-charlie-hebdo.html

[2] I. MacLeod and L. Berthiaume, "New anti-terror bill will give Canadian spies police-like powers to disrupt terrorist plots," Jan. 2015, accessed on 2015-05-20. [Online]. Available: http://news.nationalpost.com/news/canada/government-tables-anti-terror-bill

[3] B. Schneier, "Why mass surveillance can't, won't, and never has stopped a terrorist," Mar. 2015, accessed on 2015-05-10. [Online]. Available: http://digg.com/2015/why-mass-surveillance-cant-wont-and-never-has-stopped-a-terrorist

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[5] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York: Springer-Verlag, 1988, (Online version edited in 2001). [Online]. Available: http://www-ee.stanford.edu/\textasciitildegray/arp.pdf

[6] W. B. Frakes and R. A. Baeza-Yates, Eds., *Information Retrieval: Data Structures & Algorithms*. Prentice-Hall, 1992.

[7] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*, 1st ed. Springer-Verlag, 2008.

[8] K. Bankston and S. Soltani, "Tiny constables and the cost of surveillance: Making cents out of united states v. jones," pp. 335–356, Jan. 2014.

[9] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stum, "Evaluating similarity measures for emergent semantics of social tagging," in *Proc. Int. WWW Conf.* ACM, 2009, pp. 641–650.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[11] "Terrorists' personality traits indistinguishable from traits of the general population: Experts," May 2015, accessed on 2015-05-10. [Online]. Available: http://www.homelandsecuritynewswire.com/dr20150512-terrorists-personality-traits-indistinguishable-from-traits-of-the-general-population-experts

[12] F. Rudmin, "Why does the nsa engage in mass surveillance of americans when it's statistically impossible for such spying to detect terrorists?" May 2006, accessed on 2015-05-09. [Online]. Available: http://www.counterpunch.org/2006/05/24/why-does-the-nsa-engage-in-mass-surveillance-of-americans-when-it-s-statistically-impossible-for-such-spying-to-detect-terrorists

[13] M. Hartley, "Cyber threats: Information vs. intelligence," Oct. 2014, accessed on 2015-05-09. [Online]. Available: http://www.darkreading.com/analytics/threat-intelligence/cyber-threats-information-vs-intelligence/a/d-id/1316851