Introducing Privacy-Protective Surveillance:
Achieving Privacy and Effective Counter-Terrorism

September 2013

Ann Cavoukian, Ph.D.
Information & Privacy Commissioner
Ontario, Canada

Khaled El Emam, Ph.D.
Associate Professor,
University of Ottawa
Acknowledgements

Our deepest thanks go to Dr. George Tomko, whose work inspired this paper and formed the basis of the methodology for PPS. We would also like to thank Michelle Chibba, Alex Stelanov, and David Weinkauf for their significant contributions to this paper.
# Table of Contents

Introduction ............................................. 1
1. Context and Rationale ............................. 3
2. Proposed Solution: Privacy-Protective Surveillance .......................... 7
3. Intelligent Agents ................................... 11
4. Homomorphic Encryption .......................... 14
5. Probabilistic Graphical Models ...................... 16
6. The Two Phases of PPS .............................. 17
7. Conclusion ........................................ 19

Appendix A: Technical Details of the Proposed Solution for PPS ............... 20
Introducing Privacy-Protective Surveillance: Achieving Privacy and Effective Counter-Terrorism
Ann Cavoukian¹, Khaled El Emam², et al³.

This paper introduces the concept of Privacy-Protective Surveillance (PPS) – a positive-sum, “win-win” alternative to current counter-terrorism surveillance systems – and proposes a methodology for its implementation within the framework of Privacy by Design. Section 1 discusses the current context of counter-terrorism surveillance and presents the need for an alternative solution. Section 2 introduces the objectives and functionality of PPS, showing its relevance to the private sector and distinguishing its data analytics from others. Section 3 describes the feature detection abilities of the intelligent virtual agents used by PPS. Section 4 presents homomorphic encryption within the context of PPS’s analytics. Section 5 describes PPS’s ability to contextualize information through the use of probabilistic graphical models. Section 6 presents the two main phases of PPS in terms of its development and implementation. Section 7 highlights the main issues addressed by PPS and calls for the involvement of additional stakeholders to work towards refining and implementing its proposed methodology. Lastly, because of the technical complexity involved, further details of the PPS system are presented in Appendix A.

1 Information and Privacy Commissioner, Ontario, Canada (commissioner.ipc@ipc.on.ca).
2 Associate Professor, Faculty of Medicine and the School of Electrical Engineering and Computer Science, University of Ottawa (kelemam@ehealthinformation.ca).
3 Our deepest thanks go to Dr. George Tomko, whose work inspired this paper and formed the basis of the methodology for PPS; Michelle Chibba; Alex Stoianov; and David Weinkauf.
As long as the threat of terrorism is real and the global conditions that instantiate those threats continue to exist, effective measures will be needed to counteract terrorism. At the same time, in order for a free and open society to function properly, its civil liberties must be protected. Above all, privacy, as the ability of individuals to control the collection, use, and disclosure of information about themselves – referred to at times as “informational self-determination,”4 deserves the strongest protection, since it forms the basis of many of our freedoms. Despite this essential relation of privacy to our ability to function freely in our day-to-day lives, all too often sacrifices are justified as the “price” we must pay for increased security. The most pressing example of this sacrifice is perhaps found in the current efforts of governments to protect us from terrorist attacks through the use of widespread digital surveillance, essentially foregoing the interests of privacy.

The focus on surveillance by the state has been steadily growing, leading to a crescendo in scale with this year’s revelations of the U.S. National Security Agency’s (NSA’s) activities. In two recent papers we published, “Surveillance – Then and Now,” and “A Primer on Metadata – Separating Fact from Fiction,” we chronicled the steady growth of technologies that facilitate the covert operation of many surveillance activities. While the NSA has been strongly criticized for the practices of its counter-terrorist surveillance programs, it should be noted that Canada (as well as most countries) has its own secret surveillance program that targets telephone records and Internet data. This program is operated by the Communications Security Establishment (CSE) Canada. Further, a bill which was before Canada’s Parliament last year, the now thankfully defeated Bill C-30, would have imposed similar measures on Canadian telecom service providers (TSPs). Professor Michael Geist noted that a provision in Bill C-30, Section 14(4), would have given carte blanche to government agencies to install whatever monitoring equipment they chose on TSPs’ networks, effectively creating a back door to the communications of the public, without the need for a warrant or court order. Professor Geist stated that Bill C-30 would have given the Canadian government “similar legal powers” to those of the NSA, which has the capability of installing tracking equipment inside the offices of telecom providers, and compel them to remain silent about such practices. In the Canadian context, the Commissioner noted that in giving such sweeping power to the state, “… the public risk[s] being left out of the decision-making process, and Canadians risk seeing TSPs transformed into agents of the state. This represents a significant and needless risk to a free and open society.”5 Fortunately, Bill C-30 was not passed.

---

4 This term was first used in a German constitutional ruling concerning personal information collected during Germany's 1983 census.

A series of articles in the Guardian and other newspapers revealed the sweeping scope of the NSA's counter-terrorism surveillance programs in which the NSA indiscriminately collects vast amounts of data on domestic and international Internet and telephone communications, from a variety of sources. This includes prominent Internet companies and TSPs, as well as the backbone fibre-optic cables of the Internet itself – nothing appears to be off limits. The indiscriminate nature of this collection means that not only communications related to legitimate terrorist threats are being collected, but communications entirely unrelated to terrorist activities, including those of unsuspecting, law-abiding individuals.

The collection of communications that bear no relation to terrorist activities opens up the possibility of massively surveilling individuals, with no probable cause for such surveillance, and in direct violation of their civil liberties. As Edward Snowden, ex-infrastructure analyst for the NSA and leaker of classified documents shedding light on various NSA surveillance programs recently noted, “I, sitting at my desk … [could] wiretap anyone, from you or your accountant, to a federal judge or even the president, if I had a personal email.”

The rationale often given for widespread surveillance is the “war on terror” – a war that in all likelihood, will not end in our lifetime. Under the current paradigm, the intrusions into our private lives and upon our civil liberties will no doubt escalate, to the point where the creativity and innovation arising out of privacy-based free societies will be severely affected, and in turn, will diminish prosperity. It is generally well known that our current prosperity, which includes inspiring by-products such as the arts and humanities, is a direct consequence of our freedoms. What is not well-known is the knowledge that the utility of these freedoms may only be expressed within a backdrop context of privacy. Indeed, privacy, as a backdrop, nurtures the growth of creative thoughts and actions, and is at the heart of creativity and innovation, which forms the basis for prosperity in our society. Therefore, given that the war on terror will most likely continue for some time, we must pursue public safety in a privacy-protective way. The good news is that we possess the creativity and technology necessary to achieve that end result!

As noted above, one of the main problems with the current surveillance system is that data collection covers large segments of the population, irrespective of whether they are implicated in an investigation or not. A recent article in the New York Times noted that the NSA “copies virtually all overseas messages that Americans send or receive… That could very well include innocent communications between family members expressing fears of a terror attack, or messages between an editor and a reporter covering international security issues, or the privileged conversation between a lawyer and a client who is being investigated.” In the attempt to prevent terrorist efforts, our governments, in concert with corporations, have developed what appears to be a turn-key, privacy-invasive surveillance system which, with the “flick of a switch,” could easily be turned against innocent people. This is in blatant disregard of the basic tenets of a free and open society. While analysts are only expected to monitor individuals engaged in suspected terrorist activities, the possibility is nonetheless there to conduct unwarranted surveillance on a wide range of individuals, be they potential terrorists or innocent bystanders.


Most approaches to protecting privacy while ensuring measures to counteract terrorism, seek to strike a “balance” between these two interests. This invariably means engaging in the zero-sum paradigm of giving up what is perceived to be the “less important value,” namely privacy, in favour of the “more significant value,” namely public safety (imagine a see-saw – the more one side goes up, the other side must go down). This zero-sum trade-off, however, is ultimately destructive in free and open societies. Not only is it inappropriate, it is also unnecessary. Privacy and counter-terrorism measures can indeed co-exist, with both values being achieved instead of being positioned as opposing forces. Recently, CIA Director John Brennan himself echoed this sentiment when he stated that rather than viewing privacy and security as conflicting goals, the administration aims “to optimize” both.8

We can replace the zero-sum balancing act by embedding privacy into the design and architecture of surveillance systems, right from the outset. By doing so, it will be possible to accommodate all legitimate interests and objectives in tandem – in a doubly-enabling, positive-sum manner, where both privacy and effective counter-terrorism may co-exist, without diminishing the intelligence gathering abilities of the systems involved.

---

8 Jennifer Peltz. (2013, August 9). “3 top US officials discuss cybersecurity at NYC event, say they take privacy issues to heart.” The Canadian Press.
Privacy-Protective Surveillance (PPS) is a positive-sum, “win-win” alternative to current counter-terrorism surveillance systems. It incorporates two primary objectives in its design: 1) the ability to scan the Web and related databases using a “blind-sight” procedure to detect digital evidence relating to potentially suspicious criminal/terrorist activity by some, without infringing on the privacy of other unrelated individuals; 2) a technological infrastructure, based on artificial intelligence (AI) and modern cryptography, to ensure that (a) any personally identifying information (PII) on any unrelated individuals is not collected and (b) in transactions associated with targeted activity, PII and the metadata of additional “multi hop” connections will be encrypted upon collection, analyzed securely and effectively, and only divulged to the appropriate authorities with judicial authorization (a warrant).

This paper will present a high-level overview of PPS to illustrate the organizing methodology behind it, realizing that we have not delved into many of the intricacies required to develop such a system. No doubt, the “devil is in the detail,” but that detail is no different than implementing current cryptographic systems and protocols, as well as feature detection in virtual agents using machine learning techniques. What we propose is the marriage of the two, which will lead to a positive-sum execution of privacy and security, without compromising either one. This paper is meant to be the first step of a larger process, requiring the participation of experts in the intelligence community, technologists and researchers in the AI and cryptographic community, TSPs, and Internet companies. Through the release of this paper, we hope to “plant the seed” of a “win-win” solution, and highlight certain technical approaches that can implement the PPS vision.

For the sake of this paper we will be making some, possibly highly simplified, assumptions about the types of analyses that would be performed to illustrate the concepts behind PPS in a succinct manner. More sophisticated analyses can be built on top of these principles.

As illustrated in Figure 1, PPS consists of three main technological components: 1) intelligent virtual agents designed to search for suspicious activities online or in transactional databases, and once detected, flag that activity, while encrypting any personal information associated with those activities; 2) a system utilizing Secure Multi-Party Computation methods, such as “homomorphic encryption,” to allow for the interrogation and analysis of the encrypted data, to determine the potential links between various activities and individual(s); and 3) probabilistic graphical models, such as Bayesian networks, to perform inferential analysis on the anonymous data in order to calculate the probability of a terrorist threat, given the evidence revealed from the linked suspicious activities. This probability
will then be used for the purpose of soliciting a warrant to decrypt the personally identifiable files in order to allow for further investigations to be carried out.

It should be noted that a group of technologies called “encrypted search” has long been of interest to the intelligence community. From 2007–2011, the Intelligence Advanced Research Projects Activity (IARPA) funded several projects on “encrypted search,” such as the Automatic Privacy Protection (APP) program and the Security and Privacy Assurance Research (SPAR) program. Their primary objective was to permit the NSA to search a company’s databases without disclosing the search to the company or needing a local copy of the database. This is called “Private Information Retrieval (PIR)” in modern cryptography, which also forms part of our proposal: at the final stage of intelligence gathering, when a warrant is issued, the contextual information of any relevant terrorist-related activity will be retrieved from the intelligent agents through a more scalable PIR protocol. It is important to note, however, that as a standalone measure, PIR provides very limited privacy protection. Although a company would not know which personal information was being searched for in its database, intelligence personnel

11 See Appendix A for more information on PIR protocols.
could eventually obtain access to all the personal information in that database. Our PPS proposal provides much stronger privacy protection by keeping the search in the encrypted domain, up to the point where a warrant is issued (after which, of course, normal investigatory procedures would apply).

This additional level of privacy protection should be of special interest to TSPs and Internet companies such as Google, Facebook, Yahoo, and others, whose reputations are on the line. Public perception sees them, rightly or wrongly, as being all too willing to hand over electronic copies of databases of customer information to government agencies. Indeed, that companies allegedly receive large amounts of monetary compensation from the government for “cooperating” with its requests does not help the image of TSPs and Internet companies in the eyes of their customers.12 This perception, however, may change with PPS. Not only does PPS limit its collection of data to “significant” transactions or events that are believed to be associated with terrorist-related activities, it analyzes that data wholly within the encrypted domain, thus providing additional assurance to customers that, in the off chance that their PII was collected by the system, no “prying eyes” would be able to record or monitor their actions from within the system. The long-term goal of PPS is to enable companies, as a public service to help counteract terrorism, to scan their own databases, using PPS agents, and then turn over to law enforcement a copy of the encrypted files for anonymized analysis (as mentioned below) in a privacy protective manner.

Finally, it is important to note that PPS does not use pattern-based data mining for its analytics, where vast troves of data are collected and analyzed for previously unknown patterns and relationships, in the hopes of predicting the future behavior of individuals. While this technique may be useful for commercial practices such as direct marketing to customers, the potential of a high rate of false positives makes its application questionable in highly sensitive areas such as counter-terrorism, where threats to civil liberties are often at issue.13 Indeed, a 2008 study by the U.S. National Research Council concludes that “Automated terrorist identification through [pattern-based] data mining […] is neither feasible as an objective nor desirable as a goal of technology developments.”14 In addition, while the discovery of new patterns and relationships through predictive data mining techniques may provide more information, more information by itself is not helpful to intelligence analysts. What is needed, rather, is intelligent information – information selectively gathered and placed into an appropriate context to produce actual knowledge. Simply discovering more patterns and relationships without an understanding of their context only overburdens already strapped resources and increases the likelihood of infringing upon civil liberties. As Professor Fred H. Cate recently noted, this practice may in fact hinder, rather than help, counter-terrorism intelligence: “If the government failed to prevent the terrorist attacks of 9/11 because of its inability to ‘connect the dots,’ just continuing to add more dots isn’t going to help further secure our nation.”15 The information yielded through counter-terrorism efforts must have value, and in order for it to add value, it must reflect context.

---


In comparison to pattern-based data mining, the type of analytics at work within PPS may be likened to “old-fashioned police work” or what is referred to as “subject-based” data analysis. Through the use of intelligent virtual agents, PPS will be able to selectively gather information that is relevant to terrorist-related activities. Through the use of probabilistic graphical models, PPS will be able to place that information into an appropriate context, to produce useful knowledge about potential terrorist-related threats. When implemented properly, the PPS system should not attempt to “predict” terrorist attacks, nor should it attempt to “profile” terrorists. Rather, it should produce intelligent information on terrorist-related activities in a privacy-protective manner, up to the point where a full investigation is warranted by the appropriate authorities. Through this methodology, the goal of PPS would be a reduction of false positives, and an increase in the efficiency of counter-terrorism surveillance.

16 See Jonas and Harper, op. cit.; National Research Council, op. cit.
The rationale behind PPS rests on the following question: What if the intelligence gathering of a surveillance system could be done in a manner that would at least achieve the objectives of current practices, without compromising the privacy of law-abiding individuals? The methodology that is being proposed is called Privacy-Protective Surveillance.

A probabilistic graphical model (PGM) will be structured beforehand by intelligence experts comprising: (1) all of the features of interest in determining potential terrorist activity (treated as nodes in the model); and (2) the connections between these features. The graphical model will highlight features that have to be detected by virtual agents. Once developed, the agents’ task of determining what actual features were triggered by an individual will serve to prune the PGM into semantics, so to speak, so that conditional probabilities for each feature/node can be assigned and then treated as evidence to infer the probability of terrorist activity given the detected features. It is this probability inference that will, in part, be used by the court to decide whether or not to issue the encryption key to decipher the identity of the individual in question.

An array of artificial intelligent agents would be created to search for specific “features” of terrorist-related activities, in both structured and unstructured datasets. For example, among the features to be searched could be purchasing fertilizer capable of bomb-making, accessing a bomb-making website, transferring money to a “watch-listed” organization, making or receiving a telephone call from a “watch-listed” individual, and so forth. One of the most attractive elements of such a system is the fact that these agents would effectively be “blind” to “seeing” any other information they may run across during their searches. Since each intelligent agent would only be configured to search for a single “feature of interest,” it would be “blind” to everything else – the agent would be oblivious to any other “non-features” such as additional personal information. Analogous to “feature detection,” it would be triggered only by the appropriate “pre-programmed” activity. The use of such agents would thus remove the possibility of unwarranted surveillance of individuals, without weakening the intelligence gathering abilities of the system – a win/win solution, which forms the basis of our proposed PPS system.

However, these intelligent agents, each developed to search for different, specific features, would also access the associated individual’s identity, location, time and context of the activity. In order to become useful information, data must be placed in a relevant context. This context is the frame of reference from which appropriate meaning may be derived. Therefore, it is important that these intelligent agents
acquire substantive knowledge about the topic surrounding the particular feature they are designed to
detect, in order to serve as the appropriate frame of reference. The ability to place the triggering of a
feature within an appropriate context, will not only improve future assessment as to the probability of
a terrorist event, but will also mitigate an additional privacy burden.

If such a database of personal information were created in “plaintext,” it would bring with it many
privacy concerns. Furthermore, the detection of a single feature may not, in and of itself, be sufficient to
justify the disclosure of an individual's personal information to the authorities on the grounds
of counter-terrorism. An obvious example would relate to the personal information
of a farmer, who is required to buy a large
amount of fertilizer for non-terrorist-related purposes associated with farming. Clearly this
information should not be disclosed to the authorities, despite the fact that the “purchasing fertilizer” feature embedded in the agent would have been triggered.

With PPS the associated individual’s identity,
location, time and context of the activity would
be encrypted through the use of a public key,
controlled by the courts, without retaining
a plaintext version. This information would only exist within the “space of cypher text,”
and it is here that analysis and linking of the
data and the resulting inferred probability and
decisions would be made as to whether these
activities should be brought to the attention
of the court. If so, then an order would be
issued to obtain a warrant to decrypt the PII
associated with the detected activities.

Through a further development of this
methodology, the government would not
need to collect and store copies of private or
corporate databases of individuals’ personal
information. Each internet company or
TSP could tailor their intelligent agents and
features to search through their own databases.
Moreover, these agents could be outsourced to
third parties that could be granted access to
the foregoing databases. The agents would only

---

17 See Section 6 below.
store the fact that a feature was triggered, along with the encrypted personal information, which could then only be decrypted through the court’s private key. These anonymous files could, in turn, be passed along to the appropriate authorities for further analysis, as outlined below. As a precaution to prevent terrorist hackers from determining the nature and type of the feature detectors being used, those that may be sensitive to disclosure would be developed using technologies such as a PIR protocol or secure neural networks, where one could not determine the feature from the structure and connection weights alone. Further, each intelligent agent would require a “key” that could be hardware-protected, in order to operate. As a result, merely acquiring the code for an agent with a sample database would not be sufficient to determine which feature it was programmed to search for.

Homomorphic encryption refers to a cryptographic algorithm that allows computing functions to take place entirely within an encrypted domain, without any decryption of the data. With the use of homomorphic encryption, one is able to process ciphertext without the need to access any of the plaintext referent. Such processing could take the form of a query to an encrypted database of individual identities to find matching records, for example, to determine whether the same individual (as yet unknown because of the encrypted values), had triggered a number of different features. While searching through Internet and telephone communications is essential to counter-terrorism intelligence-gathering, it is but the first step of a much larger process. Once Internet and telephone communications have been thoroughly searched, any detected instances of terrorist-related activities (features) must be linkable to each other in the form of an activities graph, in order to build up a database of intelligence, for subsequent analysis.

This is enabled through the use of homomorphic encryption in which query-type processes may be directed to the database of encrypted values, in order to structure “anonymous” graphical networks of activities.
of single or groups of suspected individuals. This is necessary in order to create a database of useful intelligence, while ensuring that the personal information of individuals not suspected of terrorist activities is not accessed. The various pieces of personal information associated with the features detected would be encrypted and then stored in the database in such a way that they could only be linked together via querying the encrypted information (identity, location, time, and context).
Intelligence gathering, analysis, and decision-making are carried out under conditions of uncertainty. The goal is to reduce that uncertainty without increasing the incidence of false negatives. From a privacy perspective, however, it is the false positives that give cause for the greatest concern – innocent individuals whose privacy, and at times, entire lives, may be abruptly disrupted.

The proposal to minimize false positives incorporates the following steps: The database of counter-terrorist intelligence will be created in a privacy-protective manner, using intelligent agents to flag various activities, with all personal data encrypted; homomorphic encryption will then be used to query the encrypted files in order to collate triggered features, by individuals. Decisions must then be made as to whether a particular set of linked instances of different features (referred to as nodes below) crosses the threshold for a terrorist threat. There are many methods that may be used for such automated data analysis; for example, Bayesian networks are an extremely powerful way to represent the probabilistic relationships among the nodes in order to perform inferential analysis, for the purpose of reducing uncertainty. In order to facilitate such analysis, conditional dependencies between various features and terrorist-related threats will be determined and graphed in the form of a probabilistic graphical model (PGM).

First, the syntax of a PGM is structured beforehand by intelligence experts comprising: (1) all of the features of interest in determining potential terrorist activity treated as nodes in the model; and (2) the connections between these features structured in a “parent/child” relationship. The graphical model will highlight the feature detectors that will be developed. For example, at the top of the graph will be the node, “this-is-a-terrorist-threat,” and flowing downward from this node will be all the activities, structured as features, which if triggered, could serve as evidence to infer the probability of “this-is-a-terrorist-threat.” Once developed, the agents’ task of determining what actual features were triggered by an individual will serve to prune this graph into semantics, so to speak, so that conditional probabilities for each feature/node may be assigned based on the context of the activity, and then treated as evidence to infer (compute) the probability of “this-is-a-terrorist-threat,” given the evidence of detected features.

If the probability of a terrorist threat surpasses a pre-determined threshold, the private key paired to the public key used to encrypt the suspicious set of linked instances of features will be released by a judge through a court order and the identifiable personal information related to those instances will be decrypted for further investigation by the appropriate authorities.
An important characteristic of the proposed PPS architecture is that, due to the computational and communication burden of alternative architectures, the main repository of encrypted data is maintained by the party conducting surveillance. As noted above, however, a consequence of PPS’s use of intelligent agents is that government agencies would not need to collect and store copies of private or corporate databases of individuals’ personal information, since Internet companies and TSPs could run intelligent agents locally and tailor them to search through their own databases as well as communicate with the main repository of encrypted data as needed. While such a decentralized network of agents is the ultimate goal of PPS, such a distributed system would impose significant computational burdens on TSPs and other service providers and would have quite slow query response performance, plus the fact nonetheless remains that government agencies currently have stored access to a vast amounts of Internet and telephone communications. Because of this, we foresee the development and implementation of PPS as having two main phases.

Phase I would be limited in terms of its actors to government agencies. In this phase, government agencies would develop a “centralized” implementation of PPS, running all agents themselves, who would search existing government databases. While not ideal, this phase addresses a number of critical concerns in a reasonable way and can be implemented in the short term. Phase II would increase the scope of actors to include Internet companies and TSPs. In this phase, the network of agents would become decentralized, running not only on behalf of government agencies but Internet companies and TSPs as well. As a result, government agencies would no longer collect or store copies of private or corporate databases of individuals’ personal information, since searching these sources would be the responsibility of agents run by Internet companies and TSPs. In this way, the bridge built in Phase I would be crossed, resulting in a doubly-enabling solution where all legitimate interests and objectives are accommodated, without diminishing the intelligence gathering abilities of the systems involved.

In Phase II companies would be active participants in the surveillance enterprise. While additional controls to protect privacy would be employed as we discussed, it remains to be seen whether there is societal appetite for such participation.
The current need for effective measures to counteract terrorism, combined with the recent revelations concerning the state’s indiscriminate collection of vast amounts of data on domestic and international communications oblige us, as members of a free and open society to ask questions – not only to question the legitimacy of these counter-terrorism measures, but to attempt to redesign them in a way that respects the privacy of law-abiding individuals.

In this paper, we introduced a positive-sum, “win-win” alternative to current counter-terrorism surveillance systems called Privacy-Protective Surveillance (PPS). An extension of AI, embedding privacy directly into its design and architecture allows PPS to achieve multiple gains. Through the use of such technologies as intelligent virtual agents, homomorphic encryption, and probabilistic graphical models, PPS allows for privacy and effective counter-terrorism to co-exist in tandem, without diminishing the intelligence-gathering abilities of the systems involved. Specifically, PPS offers the development of a new system design of privacy-protective feature detection. It has the ability to scan the Web and related databases to detect digital evidence relating to potential terrorist activity, while ensuring that any PII on unrelated, law-abiding individuals is not collected. In those cases associated with targeted activity, PII will automatically be encrypted upon collection, analyzed securely and effectively within the “space of cipher text,” and later only divulged to the appropriate authorities with judicial authorization (a warrant).

While we did not delve into many of the intricacies required to develop such a Privacy-Protective Surveillance system, the detail required would be no different than implementing current cryptographic systems and protocols, as well as feature detection in agents using machine learning techniques. Our primary purpose was to illustrate the organizing methodology behind PPS in order to show that, contrary to appearances, it is possible to have both privacy and effective counter-terrorism – indeed, that we possess the technology and can develop the system design to achieve this doubly-enabling end result. By doing so, we will be able to implement strong counter-terrorism measures, while ensuring the future of freedom and liberty – a win/win proposition!
Appendix A: Technical Details of the Proposed Solution for PPS

The high level diagram of the proposed system is shown in Figure 2.

An intelligent Agent searches for a specific feature, $FT_1$ (e.g., “someone bought a large amount of fertilizer”). Once detected, it triggers the search for all the information available about this particular person. The information is separated into PII, which forms the fields of extended ID (see below), and contextual (largely unstructured) data, $S_i$. The Agent receives a public encryption key, $K_p$, from a Trusted Judicial Authority (TJA) and a homomorphic public key, $K_{ph}$, from a Homomorphic Key Holder (HKH). The Agent can also have its own hardware-protected key(s), $K_d$. The Agent encrypts the contextual information $S_i$ with $K_p$ and/or $K_d$ and stores it locally. It can be released only upon receiving a warrant/court order. The structured extended ID, on the other hand, is encrypted with $K_{ph}$ and is sent to the Search Engine Database, where it will be homomorphically searched against entries from other intelligent Agents.

![Feature Detection and Encryption Methodology Diagram](image-url)
Extended ID

Each Feature Detector/Agent searches for the following information related to a specific feature, FT_1, FT_2, ... : ID and contextual information, S_i (S_i also contains ID). To match IDs using partially homomorphic encryption, the ID is stored as an extended ID and should be a structured data set (i.e. in accordance with a pre-determined template) containing all PII that is searchable with homomorphic encryption. The extended ID may contain the following fields: name, gender, DOB, SSN, address, telephone, IP address, email address, etc. Not all the fields are necessarily filled. Each field, in turn, is structured as a set of subfields to accommodate for possible errors/variations/incomplete data. These subfields will be encrypted separately with a homomorphic public key, K_{ph}, but with different random numbers, r: E_{K_{ph}}(subfield, r).

In the case of multi-hop analyses, the agent may collect and encrypt the extended IDs of people the target ID is connected with. This will create an encrypted hierarchical tree structure of IDs, which will assist with subsequent ID matching.

Contextual information, S_i

These are mostly unstructured data containing context, time, place of an event, the extended IDs, and also PII that cannot be searched by homomorphic encryption. The examples of such PII would be a photo or other biometric information. As soon as homomorphic encryption becomes more sophisticated (e.g., a fully homomorphic encryption solution is implemented), this PII may be added to the extended IDs.

Each S_i is encrypted and stored at the agent’s site. The agent may have a tamper-proof secure hardware module or a trustworthy computing hardware platform where the agent’s internal keys, K_d, are stored. Also, the secure module can be programmed to perform some computations. S_i can be encrypted both with the agent’s internal key and with the public key, K_p, from the judicial authority. This should be done in accordance with a Private Information Retrieval/Oblivious Transfer Protocol.

 Trusted judicial authority (TJA) 

This authority would be under judicial supervision, and would generate a pair of public, K_p, and private, K_r, keys. The public keys are then distributed to the agents to encrypt S_i. The private keys are securely stored at TJA. Also, TJA establishes a Private Information Retrieval/Oblivious Transfer protocol within each agent.

 Homomorphic key holder (HKH) 

The public, K_{ph}, and private, K_{sh}, homomorphic keys are generated by a trusted homomorphic key holder, HKH. The public key is sent to all agents; the private key is securely stored at HKH. HKH is not expected to learn any actual IDs.\(^\text{19}\) There can be only one pair of public-private homomorphic keys for all agents. Note that more sophisticated theoretical solutions using homomorphic encryption or Secure Multiparty Computation also exist.\(^\text{20}\) Our proposal should be viewed as a practical example, used to illustrate the concept.

\(^{19}\) At this point, it is premature to specify who will have a jurisdiction over HKH. In the simplest scenario, TJA would also perform the functions of HKH.

Homomorphic matching of IDs

The high level diagram of homomorphic matching is shown in Figure 3.

Two agents can match their IDs in pairs using a matching protocol\(^{21}\) or other disclose-if-equal (DIE) protocol. The agents communicate with HKH that will give them the answer if two subfields coincide. Depending on the number of coinciding subfields and other rules, a probabilistic score for each field is generated, followed by a probabilistic score for the pair of two IDs being obtained. The same process is repeated for all the agent pairs and all extended IDs. There are effective ways to scale this protocol for very large data sets and for multiple parties.

El Emam and Earle protocol for de-duplication of databases of medical records

The following is a secure multi-party computation protocol for record matching and de-duplication of databases of medical records. The description below specifies the basic protocol. Implementation considerations to scale this to large data sets using a hashing scheme and more efficient homomorphic encryption, as well as taking advantage of the inherent parallelism in this protocol, have been provided elsewhere.\textsuperscript{21}

Suppose there are two parties, A and B, which hold the values that need to be compared. A wants to know if a record $x$ from A matches with any record $y$ from B by comparing specific values in these records. The records to be searched are encrypted with a cryptosystem such as Paillier (although other, and possibly more efficient, cryptosystems may be used) and the search is done in the encrypted domain, i.e. without actually decrypting the records. For that, a semi-trusted third party is required The Homomorphic Key Holder (HKH). HKH is not supposed to learn the contents of any actual records; its job is to let A know if the record $x$ from A matches with any record $y$ from B.

- HKH generates a public and private pair of keys. The public key is sent to both A and B; the private key is securely stored at HKH.
- Party A encrypts a value $x$ with the public key and also adds some random number, $r$, to obtain an encrypted value: $E(x, r)$. Using the random number $r$ is a nice property of most homomorphic cryptosystem (including Paillier) which makes it probabilistic and “semantically” secure: each new encryption will have a different $r$. This means that an attacker cannot create a lookup table for all possible values and then use a reverse lookup (“dictionary attack”) to decrypt them.
- Party B encrypts all its records values with the same public key but with different random numbers, $r'$: $E(y, r')$.
- $E(x, r)$ is sent from A to B.
- B computes the following for each of its values: $(E(x, r) \times (E(y, r')^{-1})^{r''}$ . Here yet another random number, $r''$, is generated separately for each value; the result of subtraction of two encrypted values is raised to the power of $r''$.
- $(E(x, r) \times (E(y, r')^{-1})^{r''}$ for all $y$ are sent to HKH.
- HKH, having the private key, decrypts the following: $(x - y) \times r''$. Neither $r$ or $r'$ are needed for decryption.
- If the result is zero for any record $y$, i.e. $x = y$, HKH reports to A that this particular $y$ coincides with $x$.

Note that HKH does not learn $x$ or $y$, or even $(x - y)$, from $(x - y)\times r''$. Party B does not learn anything at all before the final identification. Party A does not learn about other records in B that are not the same as $x$. 

This protocol is suitable for exact matching. There are also other protocols for inexact string matching, for example, if the name was spelled slightly differently in A and B. For that, several variants accommodating different spellings would need to be searched. Since the Paillier cryptosystem is semantically secure, it is possible to separately encrypt smaller subfields of a record, or even each character. If only some, but not all, subfields coincide, the matcher will return a probabilistic score for the matching of two records. For our purposes, these scores may be entered into the Bayesian network.

Probabilistic Graphical Models

The probabilistic scores enter a separate Bayesian network which will output the list of top matches. Each output list will contain a set of homomorphically encrypted IDs that would likely belong to the same person, alongside with the probability of such an event. These values would be inputted into the main Bayesian network of activity notes, with the results presented to a judge to decide whether or not to issue a warrant/court order.

The use of a Bayesian network allows for easily updating the results as more information becomes available.

Information retrieval

After a warrant/court order is issued, the TJA retrieves contextual information, Si, from all the relevant agents. This can be done through a Private Information Retrieval (PIR) protocol. It allows retrieval of an item from the agent without revealing to the agent which item is retrieved. There is also 1:n Oblivious Transfer protocol (also called Symmetric PIR (SPIR)), where the user (TJA in this case) does not obtain information about other items stored with the agent. Well-known PIR protocols are Lipmaa and Gentry-Ramzan. It should be noted that the communication complexity of the PIR protocols may create scalability problems. However, customized protocols have been developed and deployed on a large scale that appear to resolve some of the scalability issues with PIR protocols. For example, as mentioned above, IARPA funded the Automatic Privacy Protection (APP) program where several symmetric PIR schemes were tested. Under these schemes, the security guarantees of SPIR were somewhat relaxed in return for simultaneous efficiency in communication and computation. The test results were very promising. Also, there are works on PIR using secure hardware that achieve much better communication complexity than purely cryptographic PIR (see, for example, Wang et al. or Yu et al.). In this case, the secure hardware not only stores the encryption keys but actively participates in the computations. In order to scale more effectively, PPS would use a customized PIR protocol.

References:

23 We use Bayesian networks as an example. Other probabilistic graphical models can be applied.