High Stakes: Designing a Privacy Preserving Registry

Alexei Czeskis and Jacob Appelbaum

University of Washington, Seatle, WA, USA
{aczeskis,ssladmin}@uw.edu

Abstract. This paper details our experience designing a privacy preserving medical marijuana registry. In this paper, we make four key contributions. First, through direct and indirect interaction with multiple stakeholders like the ACLU of Washington, law enforcement, the Cannabis Defense Coalition, state legislators, lawyers, and many others, we describe a number of intersting technical and socially-imposed challenges for building medical registries. Second, we identify a new class of registries called *unidirectional, non-identifying* (UDNI) registries. Third, we use the UDNI concept to propose holistic design for a medical marijuana registry that leverages elements of a central database, but physically distributes proof-of-enrollment capability to persons enrolled in the registry. This design meets all of our goals and stands up in the face of a tough threat model. Finally, we detail our experience in transforming a technical design into an actual legislative bill.

1 Introduction

Washington State, like fifteen other US states and the District of Columbia, has legalized marijuana for medical use [1]. However, Washington State is the only one that does not yet have a medical marijuana registry [16]. This paper details our experiences in helping multiple stakeholders design a legal framework and the technology behind a privacy preserving medical marijuana registry. Additionally, we believe our design to be broadly applicable for many other kinds of registries.

We began by directly and indirectly gathering information from multiple stakeholders like the ACLU of Washington, law enforcement, the Cannabis Defense Coalition, state legislators, lawyers, and many others. Each group had their own goals and agendas, which often conflicted with the goals and agendas of other groups. These interactions, generated many complex design goals, technically and socially imposed challenges, among which was the need to function in the face of a very strong adversary.

As a result, the exercise drove us to study a new class of databases or *registries* that we believe have not previously been discussed in literature or deployed in the wild. Specifically, our proposed registry design does not store any Personally Identifiable Information (PII) – either in digest or encrypted form. Instead, we delegate limited information out to *proof-tokens*, which are given to enrollees (people enrolled in the registry). Enrollees can use the proof-token to prove their enrollment in the registry. Additionally, because it is impossible to indentify enrollees by having access to the registry, enrollees can deny that they're enrolled by hiding or destroying the proof-token.

We begin by giving a background of registries. Next, we motivate the need for a new type of registry – a unidirectional, non-identifying (UDNI) registry. We then outline the goals and challenges for a successful UDNI registry design. Next, we provide several example architecture designs and explain why they fail to meet all of the required UDNI goals and break in the face of our threat model. We then present our proposed design and examine it in the context of a detailed case study that covers each aspect of our design in depth. The case study focuses on the proposed medical marijuana registry in Washington State and is grounded in actual facts and concrete discussions. Finally, we discuss what it means to put this type of technology into law, give some pointers on careful implementation, and finish by examining a couple of other interesting, relevant topics.

2 Background

Most modern societies maintain records about people – who they are, where they live, what they are allowed and not allowed to do. These records often manifest in the form of databases or *registries*[1] and are often crucial to how certain

[1] Note that the terms *database* and *registry* are often interchangeable. However, *registry* is often used to refer to a holistic system (including people that use it); this is why we default to using this term.

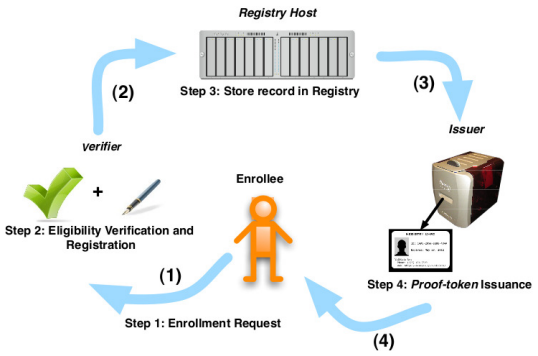

Fig. 1. Generic enrollment process

Fig. 2. Proof of enrollment

aspects of law and order are enforced. For example, in many countries people legally drive only if the state driver's licence database says so and every driver must carry proof in the form of a driver's license while operating a vehicle. As another example, entries in common medical prescription systems tell pharmacies which pharmaceuticals to fill, to whom, and when. As a final example, infectious disease registries record appearances of certain diseases and may be used to detect potentially dangerous outbreaks and epidemics.

Some registries claim to be purely statistical, privacy preserving, or even anonymous. They try to achieve this goal by utilizing a variety of well-studied techniques like k-Anonymity [19] or Differential Privacy [9]. Indeed, the "privacy in databases" community is quite rich with literature that provides models and metrics for achieving some *reasonable* level of anonymity for entries in a database. The majority of these techniques give a heuristic for how much PII a database maintainer must trim (or how much noise they should add) in order to get a set of data that is lean to a point where a person can be mapped into a large set of entries (instead of just one). Other techniques deal with how many and what kind of aggregate queries should be permitted on a PII database in order to maintain anonymity and privacy for any registered person's data. Dwork gives a good survey of database privacy techniques in [10]. Nevertheless, attacks against these metrics do occur [15], and statistical registries can morph into identifying registries overnight.

However, a portion of registries (currently implemented as identifying registries) only require a uni-directional link to function – that is, the registry's sole use is for people to prove that they are "enrolled" in the registry. These registries are not designed to enumerate members or to store any information about the enrolled members other than the fact that they are enrolled. We call such registries *uni-directional non-identifying registries* or UDNI registries for short.

TERMS AND DEFINITIONS

Before diving deeper into the registry sea, we first present some common terminology. A *verifier* verifier checks whether a person is entitled to be listed in the registry. If so, the verifier registers or enrolls the person in the registry, which is stored by the *registry-host*. Once enrolled, the person is referred to as an *enrollee*. If supported by the registry, an *issuer* provides the *enrollee* with proof-of-enrollment, called *proof-token*. While proof-tokens can come in a variety of form factors, we assume that it will be physically manifested in the form of an ID card. Figure 1 demonstrates the enrollment process. Post enrollment, an *enforcer* can ask an individual to provide the proof-token or face the legal consequences of not being enrolled. This process is illustrated in Figure 2.

3 Motivation and Goals

In this section, we'll discuss examples of registries that could be made to be UDNI, what challenges arise with providing security along with functionality and privacy, and outline some of the desired goals for a privacy preserving UDNI registry along with a threat model it must protect against.

3.1 Motivating Examples

We provide two brief motivating examples of how UDNI registries can be used and where designs can fail.

EXAMPLE 1

Bob has finished his undergraduate degree in political science and has just been accepted to law school. Unfortunately, Bob has also been diagnosed with cancer. Bob begins chemotherapy (chemo), but the chemo-induced nausea and vomiting make it difficult for Bob to study. Since Bob doesn't seem to respond to standard antiemetic drugs, Bob's doctor suggests that Bob try Medical Marijuana, as some studies have shown it to work in such cases [17]. Bob is afraid of being arrested – he has seen news articles about patients who are mistakenly arrested by police and are only able to prove their innocence much later. Bob's doctor recommends that Bob register with the state's Medical Marijuana Registry (MMR), which will issue Bob a card that he can carry in his wallet and present to police in case there is ever a question. Bob's doctor says this registry is also private and secure. Bob agrees to try Medical Marijuana and follows the doctor's advice to join the registry.

A couple years later, Bob's cancer is gone, his career has taken off, and he decides to run for public office. Based on an anonymous tip, the opposition hacks the server that hosts the database for the state's MMR and releases the database anonymously on-line. The opposition then issues an ad saying that Bob is a drug addict and is probably currently using other drugs. Bob is shocked and tries to explain to the voters that he was using marijuana by his doctor's recommendation, but the opposition's tactic of shock and awe have won – before the voters are able to logically think through all of the facts, the election is over and Bob has lost.

DISCUSSION

In this example, Bob followed his doctor's suggestion and used a controversial, recommendation-only medicine in order to stem his chemotherapy induced symptoms. Bob also registered in the state-provided registry in order to receive a state-issued card that would protect him from arrest if he were ever stopped with possession of the medicine. Unfortunately, the state registry stored enough information so that when the registry was compromised the attackers were able to uncover Bob's name from the registry and cause irreparable harm to his reputation.

EXAMPLE 2

Sue loves the wilderness, especially fishing. Her sister Mary, however, finds fishing and hunting distasteful – so much so, in fact, that Mary and Sue have had numerous arguments about this issue. Sue likes to hunt and fish, but she doesn't want to feud with Mary either. Sue decides to try to keep mum this season and avoid confrontation with Mary. This strategy seemed to work so well, that when Mary one day did ask Sue whether she still hunted and fished, Sue automatically said "no".

Sue decided to go fishing one last time and had great success – she caught a huge fish. On the way back, however, she was stopped by the local park ranger and asked to present her fishing license. Sue's fishing license had her name on it, which the ranger noted and included in his daily report. The ranger submitted his report at the end of the day to the office assistant, Mark. As Mark typed up the report, he noted Sue's name. Mark was friends with Mary and he thought it would be amusing to let her know that he had come across Sue's name in the papers. Mary was furious; not only had Sue continued to fish and hunt, but she also lied to Mary.

DISCUSSION

In this example, Sue wanted to keep her hobby private, but she had to enroll in a state registry because her hobby required a state issued license. Unfortunately, the license contained Sue's name, which was recorded by a ranger during a routine check. In this manner, Sue's hobby was disclosed against her wishes; this is a kind of misclosure that could be avoided by design.

3.2 Goals and Challenges

Using these examples, we now derive goals for a UDNI registry. These goals are the result of consulting with multiple stakeholders during an actual medical cannibis bill [3] drafting process that was later signed into law[2].

FUNCTIONALITY GOALS

First and foremost, the registry must be functional. The registry should support at least the following features:

- **[G1]** *Controlled enrollment* – Only those persons that actually belong in the registry can be enrolled.

[2] Most of the registry system was line item vetoed with a suggestion for it to return in a bill by itself

- **[G2]** *Provable enrollment* – If they so desire, the enrollee can provide proof of enrollment. This "proof of enrollment" must pass police "muster" (the police must accept this as valid proof). Furthermore, no non-enrollees should be able to claim enrollment in the registry (the proof should be reasonably hard to fake or forge).
- **[G3]** *Deniable enrollment* – Enrollee may deny enrollment in the registry if they so desires.
- **[G4]** *Revocable enrollment* – An enrollee may be removed from the registry upon a valid request.
- **[G5]** *Expiring enrollment* – The registry must support the expiration of entries after a certain fixed period of enrollment. Note that this goal differs from *revocable enrollment* because the former refers to revoking an enrollment based on name, while the later refers to revoking enrollment based on enrollment date.

SOCIALLY IMPOSED GOALS

Multiple stakeholders are involved in the registry system and often, they can have conflicting goals and complicated relationships. This results in subtle, but architecturally interesting goals and tensions. In order to maximally satisfy this criteria, the UDNI registry must support:

- **[G6]** *Inexpensive implementation and maintenance* – Ultimately, the registry-host role will most likely be filled by a government organization. Furthermore, the registry itself will likely generate little or no income. Consequently, the registry must be inexpensive to implement and maintain.
- **[G7]** *No new proof-of-enrollment hardware or software* – Proof-of-enrollment should be possible without specialized hardware or software for the enforcer. Practically speaking, law enforcement is reluctant to add hardware or software to their existing tools for officers in the field. This will not only facilitate quicker adoption, but will also help satisfy `G6`.
- **[G8]** *No social stigma or unintended consequences* – The enrollment status should be non-obvious to casual onlookers (e.g., proof-token color and features should be considered in the context of other identification systems in current deployment).

SECURITY GOALS

Adding further dimensions to the design, is the possibility of a powerful attacker. We assume the attacker is able to:

- *Mount Network attacks* – The attacker can perform all known network based attacks. For example, the attacker can hijack DNS or perform man-in-the-middle attacks.
- *Steal the Registry* – We assume that the attacker will be able to steal the registry. Once database is stolen, the attacker can execute very powerful brute force attacks. Additionally, we assume the attacker knows the full domain of all possible entries in database (e.g., the attacker knows the names of people residing in a particular region).
- *Employ Social Engineering* – The attacker can threaten or socially engineer the maintainers of the registry into accessing the registry and reading back values.

Given the broad abilities attacker, we claim that the registry should have the following security goals:

- **[G9]** *Minimal PII required to enroll; destroyed after use* – The registry may require certain linkable pieces of PII, but such information must not be kept beyond the time needed to produce and distribute a proof-token.
- **[G10]** *No PII in registry* – The registry must not store any PII. This will assert that a compromise of registry does not reveal identities of enrollees.
- **[G11]** *No external identification requirements* – Enrollees must not be required to carry or produce any additional documents in order to prove enrollment in the registry. This will assert that no PII is ever transmitted during the proof-of-enrollment phase and will prevent the network attacker from gathering any useful information.
- **[G12]** *Positive verification does not produce PII* – Positive Verification should not create additional PII details. Note that a negative verification may, however, produce PII related data in the context of fines court proceedings, or other law enforcement actions.

4 Architecture

While many registry architectures are possible, very few actually meet all of our desired goals and stand up to our threat model. In many cases, it may be very subtle or seemingly unintuitive why a certain design may fail. To this end, we first discuss some promising, yet flawed architectures before finally proposing our design.

DESIGN 1: PII DATABASE

The first registry design to consider is one that stores all possibly relevant data. For example, the registry could store the names of enrollees, which entity served as the verifier and when. To prove enrollment, an enrollee could present any acceptable identification (e.g. a driver's license) to the enforcer (e.g. policeman), who would then verify enrollment by calling the registry maintainer or visiting a portal and entering the ID information. This will clearly enable some functional goals like G1, G2, G4, and perhaps G5/G6/G7, but it will fail to meet G3 in a very serious way. Socially imposed goals such as G8 and G9, G10, G11 and G12 are nearly impossible to satisfy with such a simple design.

If the database or the verification mechanism were ever compromised, then confidential enrollee PII would be obtained by the attacker[3]. Moreover, this system is subject to the whims of insiders – an employee may be coerced into disclosing sensitive data or could confirm the presence (or absence) of a specific individual in the database based on personal or financial interests [6, 7, 21].

Similarly, this class of designs encourages extremely unsafe practices such as the collection of large amounts of PII during registration and proof-of-enrollment. This creates an environment where large amounts of PII is collected, transfered, and/or stored by an unknown number of parties – any of which can record or expose it in unauthorized manners. Furthermore, this design relies on traditional ID cards, which contains a large amount of PII that is irrelevant to the registry. Finally, this type of a design carries with it a negative "big brother" social stigma.

DESIGN 1.5: SPRINKLE IN ENCRYPTION/BLINDING

A better decision would be to encrypt stored PII. One approach may be to encrypt the database using a single key. While at first glance, this may appear to help protect against an attacker who is able to "steal" the database, this design still likely fail because we assume the attacker will likely be able to gain physical access to the database machine and thus compromise the decryption key [14]. Additionally, the encrypted database will also not survive in the face of a malicious or coerced employee (who could access the entire database).

A more sophisticated approach may be to encrypt each database entry with a different key and store each key (or a password used to derive the key) on the *proof-token* that's issued to the enrollee. This approach would indeed distribute the encryption keys in such a way, that each enrollee's information would only be accessible if given access to their proof-token. In order to verify an enrollee, an enforcer could either transmit their encryption key to the registry maintainer and receive decrypted data or fetch particular encrypted data from the registry and decrypt it locally using the proof-token decryption key. In the first case, transmission of the decryption key makes it vulnerable to recording by a malicious registry employee. In the later case, the enforcer would need specialized equipment – directly violating our goal G7. Furthermore, in both cases, the enforcer could record or photocopy the proof-token (or the enrollee could lose it) – completely revealing the enrollee's data. The same logic holds for a system that relies on enrollee-remembered passwords, except with the added complexity of enrollees forgetting passwords (especially in times of distress). Extensions such as using one-time-pad encryption or re-encrypting the enrollee's data also fail (either because of complexity and cost of implementation or because of the same reasons as plain encryption).

These solutions still require the extensive collection of PII for verification and enforcement, and still carry a negative stigma of having the enrollee's data stored in a database. Although technically-savvy people often understand the protections offered by encryption, others don't and forgo the benefits of such systems because of perceived privacy concerns – we found this to be true in our conversations with people who work closely with existing registries.

DESIGN 2: HASH DATABASE

Instead of storing enrollee PII in a database, the registry could store a one-way digest of enrollee PII. For example, the registry may store a hash of the enrollee name or driver licence number. Note that the hash must also include a per-enrollee secret, otherwise a stolen database can be brute-forced by an attacker who can easily discover the domain of all possible enrollees. In any case, this class of designs faces the same problems as *Design 1.5* above: in order to verify an enrollee's, the enforcer would again require special hardware or would need to send enrollee PII to the registry maintainer for verification. Both options are unacceptable in the context of our goals and threat model.

DESIGN 3: NO DATABASE

Having run into fundamental problems using a central database, we now turn to exploring fully distributed approaches (completely lacking a central database). One approach in this space could be to issue proof-tokens containing encrypted

[3] Interestingly, as a possible feature creep some states have actually been known to offer information in such databases up for sale [12, 13].

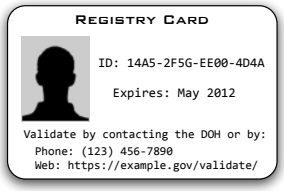

Fig. 3. A proof-token as per *Design 4*.

	G1	*G2*	*G3*	*G4*	*G5*	*G6*	*G7*	*G8*	*G9*	*G10*	*G11*	*G12*
Design 1	**Y**	**Y**	N	**Y**	M	M	M	N	N	N	N	N
Design 1.5	**Y**	**Y**	M	M	M	M	**Y**	N	N	N	N	N
Design 2	**Y**	**Y**	**Y**	**Y**	N	**Y**	N	**Y**	**Y**	**Y**	N	N
Design 3	**Y**	**Y**	**Y**	**Y**	N	**Y**	N	**Y**	**Y**	**Y**	N	N
Design 4	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**

Fig. 4. How each design meets the system goals (Y = yes, N = no, M = maybe).

data to enrollees. This could be in the form factor of a card with a hexadecimal string on it. The tokens would at the least have to encode the enrollee's identity (to prevent forgery and impersonation) and an expiration date (to support goal `G12`). In order to access the decoded data, an enforcer would need to decrypt the data. As a variation of the above, proof-tokens could contain unencoded data along with an authenticating signature. The cryptographic signature would cover all of the data on the proof-token. To verify the authenticity of the token, an enforcer would need to verify the signature of the data.

This design is attractive, but unfortunately, verifiers would be required to carry specialized cryptographic equipment. Additionally, in order to mint proof-tokes, PII would need to be collected – making this a nonviable class of designs.

DESIGN 4: UNLINKABLE TOKEN/DATABASE HYBRID

In order to eliminate the need for the enforcer to collect/transfer PII or carry additional devices, and to remove PII from the database, we propose a hybrid token/database system. In this design, enrollees will be issued a proof-token in the form of a card with their photo, a random nonce, and an expiration date (see Figure 3). The registry database will store the issued nonce and the associated expiration date – it will *not* store the associated photograph. The card will be printed with the same anti-forgery techniques (e.g., lenticular printing or watermarking) that are deployed for other government issued IDs. In order to verify the validity of an ID, an enforcer must check that it contains all of the required anti-forgery signs, that it has a valid expiration date, and that the photograph matches the enrollee in question. Note that law enforcement, park rangers, and club bouncers are accustomed to doing all of these steps already. For extra verification, and to check for revocation, the enforcer could contact the registry (either via a phone or via and verify that the nonce has not been revoked).

Note that this design does not require the database to store any PII – stored data could be made public without any negative consequences to enrollees. Also observe that the enrollee can prove their enrollment by presenting the proof-token or deny enrollment by hiding or destroying the token. The storing of the random nonce in the database allows for token revocation and the presence of the expiration date on the proof-token allows for easy expiration checking. Also note that enrollment only requires a photo (which will not be stored, but only used during the issuance phase). Finally, enforcers will not be required to carry additional equipment or collect PII to check an enrollee's proof-token.

We go further in-depth regarding this design in the following section, where we present a case-study of an actual system currently being developed.

5 Case Study

We now analyze our system as it would fare if adopted by the medical marijuana registry (MMR) currently being considered (but not yet not implemented) in Washington State.

5.1 Background and Assumptions

BACKGROUND

In the United States, fifteen states and the District of Columbia have approved marijuana for a variety of medical uses [1, 16]. However, because cannabis is not approved for medical use on a federal level, it cannot be regulated through the regular prescription system like other controlled substances such as hydrocodone or morphine. This means

that doctors cannot issue prescription for qualifying patients, patients cannot obtain medical marijuana at pharmacies, and police have difficulty determining whether a person is a criminal or a patient in pain when in possession of marijuana. However, doctors can talk to patients about medical marijuana, make recommendations and record them in patient records. The doctor can provide a copy of this recommendation to the patient, who can then use this medical documentation as part of a legal defense against prosecution in court. Nevertheless, medical marijuana patients could still be arrested and detained by police – a fairly large inconvenience to sick people (especially if they are in pain) and a matter of public record.

In order to clear the regulatory haze, fifteen out of fourteen states (and the District of Columbia) have begun to design and deploy medical marijuana registries [16]. These registries enable law enforcement a way, sometimes a quick way, to verify a patient's legal status and offer patients protection against unwarranted arrest, search, and seizure.

These registries are born into an interesting ecosystem. They enable enforcers to identify persons possessing marijuana legally under State law, but at the same time they identify people who may be breaking Federal law, which does not exempt medical use of marijuana from criminal liability. Enrollees want to be able to prove enrollment in some cases, but be able to deny it in others. Law enforcement wants easy verification of enrollment, but no additional equipment to do it. The State wants cheap and quick implementation and maintenance. Moreover, because Federal seizures of State material have occurred, the threat of physical or legal removal of the registry database [5] and arrest of enrollees is quite real.

ORGANIZATIONAL ASSUMPTIONS

We assume the proposed registry system will be run by the state Department of Health (DOH). The DOH will issue cards that stand as proof of enrollment in the registry. Additionally, the DOH will confirm doctors as eligible to make medical recommendations to patients and will keep on file the confirmed physical mailing address of the authorized doctors. We do not assume that the DOH is trustworthy in every way and we assume that they may even be subject to a subpoena or a National Security Letter (NSL) supported by a gag order that prevents them from disclosing the receipt of such a subpoena.

Note that every state DOH subject to Federal legislative action and is not willing to expose state workers to a large level of controversy; the DOH may be willing to issue privacy preserving cards but they may not wish to take on the liability of having a database of PII under their purview.

Additionally, we assume that physical mail will not be subject to constant monitoring and inspection beyond cursory recording of source/destination addresses. Specifically, we believe that the Federal government will not seize all outgoing mail for the DOH and record the contents of every letter starting at the first day of operation. We additionally assume that a doctor's office is not automatically subject extralegal action such as a document retrieving raid without due process. A doctor's records are a likely target but the actual doctor who writes a recommendation is decoupled from the registry entry after it is issued.

5.2 Registry Enrollment and Card Issuance

The enrollment phase consists of several steps. We examine each in turn below in the context of a fictional patient Robert and his physician, Jane.

STEP 1: DOCTOR-PATIENT RECOMMENDATION

Robert is an elderly man with serious pain management issues. Dr. Jane suggests that Robert enroll in the medical marijuana registry run by the State. Dr. Jane explains the system to Robert and affirms that she considers it to be a fairly safe system with only a small number of tradeoffs. Robert decides to enter the registry based on the advice of his doctor.

STEP 2: REGISTRY ENROLLMENT

Dr. Jane connects to the DOH registry website and authenticates with her account credentials as issued by the DOH. To further reduce the PII held by the Department of Health, Dr. Jane uses The Tor Browser bundle [2, 8] to access the website[4]. She requests that the DOH issue a valid card, and she submits the only required piece of information – a photograph of Robert that is to the same standard as the state driving license.

[4] This prevents local observers from noticing that she is connecting to the DOH – it also prevents the DOH from having an IP address in their logs that is meaningful after her session has ended

The whole enrollment process is shown in Figure 5. Note that it differs from the enrollment process shown in Figure 1 with the addition of using the doctor as a privacy preserving *proxy*[5].

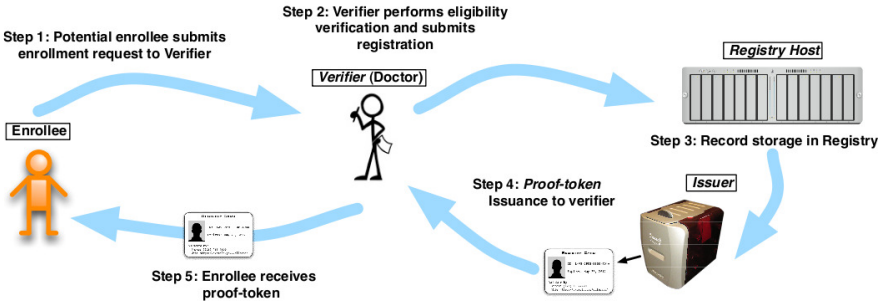

Fig. 5. Privacy preserving enrollment process

STEP 3: CARD ISSUANCE

The DOH takes care to not store this photo during or after the card production process. This is performed in the following manner:

1. The DOH computer system generates a random nonce that is unique for Dr. Jane's session and it automatically creates an expiration date one year in the future.
2. The DOH system stores the random nonce and the associated expiration date in the DOH MMR database.
3. The DOH system simultaneously prints a plastic card with Robert's image, the random nonce, and the expiration date.
4. The DOH system also prints an addressed envelope (if one hasn't already been printed that day) and a receipt that tells an operator to place the card with a particular nonce into a particular envelope (at the end of the day, the envelope is sealed and mailed to Dr. Jane's office).

Note that the above operations are performed as an atomic, blocking transaction – they either all succeed or all fail and Dr. Jane (or her nurse) must wait for the process to complete. If the process completes, the DOH effectively assures Dr. Jane that the stated registry system processes have completed as expected. Finally, Dr. Jane is presented with a receipt number, which she writes down in Robert's file. Robert does not have the protections provided by the registry until the card arrives and until that time, he has whatever protections are provided by Dr. Jane's recommendation letter.

A week passes and Robert returns to the office and meets with Dr. Jane as he would during any office visit. Dr. Jane has received the letter from the DOH and opens the envelope; inside she finds other envelopes with the appropriate receipt number for Robert. Dr. Jane hands Robert his card in private and Robert examines the envelope to notice that it is sealed and appears in an untampered state. Robert breaks the DoH seal and together with Dr. Jane they confirm that this is his valid card. For her records, Dr. Jane records the number on the card in Robert's file. The receipt number is no longer retained and Robert inspects the card. Robert inspects the card and notices important key features:

- Robert's photo under a Lenticular coating and other anti-forgery features such as a holograph of the State Seal
- An easy to read registry number with an expiration date that is set to expire in a year
- A secure (HTTPS) URL for the DOH verification website and a toll free number to call for verification
- It states that it is "State Issued Photo ID"

The card otherwise blends in with the other cards in his wallet.

[5] We did not include the *proxy* in the previous figure because it is a generic explanation of enrollment roles; the proxy is a privacy-adding, non-standard role

5.3 Enforcement and Proof-of-Enrollment

DECIDING TO DECLARE PROOF OF REGISTRY MEMBERSHIP

On the way home from the doctors office, Robert stops at a local medical marijuana dispensary in order to purchase his recommended medication. Following standard procedure, the dispensary operator challenged Robert for proof that he was a qualifying patient. While looking around the shop it became clear to Robert that he had no intention of revealing his name to complete strangers and he opted to use the privacy preserving registry card (instead of his patient records).

The dispensary operator verified the photograph visually and as a final step of the verification process, the operator launched a copy of the Tor Browser Bundle and visited the secure website for the Department of Health[6]. The dispensary operator entered the registry card number and submitted it for verification. The DOH website verified that this card was valid, not expired, and not revoked. The dispensary operator was now fully satisfied that Robert was currently eligible to purchase goods from the dispensary and warmly welcomed a new customer.

REQUESTED TO DECLARE PROOF OF REGISTRY MEMBERSHIP

On the way home, a police officer observes Robert produce a marijuana-like substance from his bag. Following standard procedure, the officer decides to check if Robert is carrying an illegal substance and prompting Robert for an explanation. Robert produces his state issued registry card. The officer verifies that Robert matched the photograph on the registry card, that the card carries all of the required anti-forgery features, and then proceeds to call the dispatch to verify the details on the card.

Robert is entirely compliant and waits while the police officer receives confirmation over the radio. While Robert has no idea if the police radio is encrypted or if the phone call from dispatch to the Department of Health is somehow secure, he feels content that none of his private information is being transmitted since he did not give the officer any PII to transmit. Meanwhile, the police officer reads off the details on the card, waits, and hears the dispatch officer respond that the card is indeed valid, non-expired, and non-revoked. The officer thanks Robert for his compliance, tells Robert that he's free to go, and wishes him well with his treatment plan.

5.4 Registry Database Compromise

THE DEPARTMENT OF HEALTH HAS A MALICIOUS INSIDER

Shortly after Robert enrolls in the registry, a DOH employee gains access to the Department of Health registry database and sends it to the local newspaper. However, the newspaper is only able to extract and print the total number of valid cards in the registry, the number of cards that were revoked and the dates of expiration.

THE DEPARTMENT OF HEALTH IS INVESTIGATED

After the high profile leak there is a surge in enrollment by many people who previously feared entering into the registry. This surge attracts the attention of the Federal government. Law enforcement agents raid the state DOH and seize the registry database[7].

The Federal agents involved are unable to extract any specific patient names. Furthermore, no state employees are interrogated or prosecuted as there is no information that could be gained from them besides what is already known. Such an activity is, by its very nature, disruptive to ongoing card issuing attempts and no further information collection is possible after service is disrupted.

5.5 Renewal and Unenrollment

RENEWAL OF REGISTRY MEMBERSHIP

Robert finds that his treatment plan has worked well for him and after one year he asks Dr. Jane to renew his enrollment status in the registry. He schedules an appointment and goes through the enrollment process with his doctor as before.

CANCELING REGISTRY MEMBERSHIP

Robert decided that he no longer needs to use medical marijuana for pain management and consequently does not

[6] The operator also made sure that the HTTPS certificate matched the well known certificate for the DOH

[7] Across the United States there are currently legal battles and law enforcement raids whereby the Federal government is attempting to force the disclosure of the list of enrolled patients

require protection provided by the registry. Even though the card will expire, Robert cuts his card in half and mails the number half of his card, without his photograph, to the DOH from a public post office. He does so without a return address.

6 Discussion

MISTAKING CRYPTOGRAPHY FOR A PANACEA

A privacy preserving registry system ensures that many privacy properties that were once a function of policy become a key part of the actual technical design. Privacy as a function of design is an absolute necessity when deploying a system in a legally hostile environment. While cryptography can often turn a policy goal into a technical reality, it's not always feasible to deploy because of confusion, cost sensitivity, lack of trust in perceived to be complex systems, and fear of serious legal or physical consequences. For example, a system without a photograph with binding to a name, with a per name secret is an entirely reasonable security system – it is also a system that only an expert can understand and is nearly impossible to deploy in a way that will not enable coercive disclosure of real names by verifying parties. Furthermore the world may someday be ready for fully anonymous credentials but the first deployments will be extremely difficult and world-shifting for law enforcement and enrollees alike. The system we propose makes a small anonymity compromise at the level of enrollee tokens by including an image. However, this compromise neither enables easy privacy violating attacks nor adds PII to a central system – it does, on the other hand, make a system that meets socially imposed restraints. We believe that this is a great improvement over the status quo.

DRAFTING LAWS FOR PRIVACY BY DESIGN REALITIES

A group or a person wishing to write a privacy preserving law would be well served to carefully and specifically phrase certain design goals in a registry creation bill. A concrete example is to ensure that the bill will not permit collection of unneeded PII and to ensure that any such data is kept in a one-way, non-reversible format. Encryption is simply not enough for many high risk registries – disclosure of cryptographic keys may be accidental or forced through any number of means (e.g., rubber hose cryptanalysis) – the stakes may simply be too high for designs that allow for both forward and reverse queries of the dataset.

Some stakeholders completely reject the concept of a registry at all costs [4] because of privacy concerns. Indeed, there may be very compelling reasons to ask if a registry is really the step that society wishes to take, especially given the concerns that a poorly designed registry may pose to otherwise lawful citizens. However, when a registry must be deployed, we believe that it is imperative to reduce the total PII to the absolute minimum level possible.

PRACTICAL IMPLEMENTATION ISSUES

The devil is often in the details and practical implementation decisions can often make or break the privacy properties of a registry. For example, in the case study we presented above, the following details need to be taken into account:

- The unique registry identifier must be chosen from a uniformly random set and must be globally unique; we assume that the token does not need to be easy for a human to remember and so Zooko's triangle [20] is not a problem for any of the parties involved.
- To slow any verification process that may become a registry identifier oracle (allowing a forger to guess valid registry identifiers), we strongly suggest only allowing queries to be performed on an ID in conjunction with an expiration date. A forger would have to guess both in order to receive a valid answer. Additional rate limiting would also help to curb abuse.
- Data retention in such a registry is an extremely important issue – while we discuss not retaining PII, including IP addresses, of the enrollee, we need to also stress the importance of erasing data as soon as it is no longer needed. By not having data, the registry prevents a "data valdez" [11] incident from occurring.
- Any database field that is unique per-person must be stored in a one-way, non-reversible manner. It may make sense to protect some non-enrollee PII data with a scheme like scrypt [18].

PRACTICAL COMPROMISES

Often in system design, and especially in security, one has to make tradeoffs and compromises. For example, one fundamental property of an anonymous registry is for the ability of some enrollees to have duplicate entries. Observe that since enrollment is deniable, there is no way to verify whether or not a particular person has already been enrolled

in a database or not. As another example, consider how revocation is impacted by providing deniable enrollment in a system. Let's observe how revocation would work in the context of our case study. Proof-token IDs can be marked as revoked – making the proof-token cards invalid. However, in order to actually revoke an ID, one has to discover which ID number to revoke. Because the system we present provides deniable enrollment, an enrollee can always hide the fact that he's enrolled and never give up his ID number – making it hard to revoke. Note that doctors may also have the enrollee's ID number, on file, but fishing for that a particular enrollee's doctor is a difficult, privacy violating, and legally murky endeavour.

Taking a step back, this is a traditional tension between privacy and security where more information causes privacy concerns for the enrollee, but more data may (or may not) provide greater security. Often a suitable security tradeoff is nevertheless possible without sacrificing much privacy – for example, we don't currently require full body scans or retina scans to have a drivers license – a close match is good enough for verification by a law enforcement officer.

LAST BITS OF ADVICE

Writing technology and privacy solutions into a bill is a challenging topic and requires meeting with people from many different groups and with a varied set of incentives. It is not possible to make every stakeholder happy as some stakeholders hold on to opinions based on idealogical grounds and are not willing to compromise.

7 Conclusion

In this paper, we discuss some of the key privacy properties offered by existing registries and discover that a fundamental space exists for unidirectional, non-identifying (UDNI) registries. Through conceptual investigation, extensive discussions with multiple stakeholders, and empirical analyses of current designs, we outline the goals for a holistic UDNI registry that takes into account not only technology, but also the people that will be using it. For example, some key goals that we meet include *provable, but also deniable enrollment* and we have *no PII stored in registry*. We also mandate that the registry not require any additional hardware or software beyond what it will take to store and operate the central registry system (i.e., the police won't have to carry any additional hardware). We complicate the design space further by supporting a sophisticated and strong adversary, who can not only interpose on all network traffic, but can also physically steal the servers on which the registry resides.

Next, we generate and systematically analyze various candidate systems and examine why and how they fall short of our goals and threat model. Interestingly, we find that complicated cryptographic techniques are insufficient to solve our problem. Instead, we propose a hybrid system that leverages elements of a central database, but physically distributes proof-of-enrollment capability to persons enrolled in the registry. This design meets all of our goals and stands up in the face of our threat model.

We explore our design further in the context of an actual case study focused around the medical marijuana registry currently being discussed in the State of Washington. Finally, we discuss how we translated our technical design into legal language, which we then helped incorporate into a Washington State bill that was recently signed into law.

This paper contributes a deep exploration of privacy and anonymity issues in certain types of registries, serves as a case study in holistic system design, and provides our experience in transforming a technical design into legalize for inclusion in a bill.

8 Acknowledgements

This publication was made possible in part by Grant Number HHS 90TR0003/01. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the HHS.

A giant thanks to everyone who knocked down our early designs, gave us their stakeholder feedback that we could not ourselves see or understand, and read our early drafts. A special thank you to Alison Holcomb, Brian Alseth, and others at the ACLU of Washington for their insight, support, and guidance. Also, a great thank you to John Gilmore for his brilliant outlook and deep understanding of all of the underlying issues. Thank you to Zooko Wilcox-O'Hearn of Least Authority Enterprises, Andy Isaacson & Leif Ryge from Noisebridge. We're especially grateful to Tadayoshi Kohno from the University of Washington, Kelly Caine from the School of Informatics and Computing at Indiana University, Dr. Nadia Heninger from Princeton University, and Phillip Mocek and Ben Livingston of the Cannabis Defense Coalition. Thank you to everyone else we forgot to mention.

Bibliography

[1] RCW 69.51A.010, Section 4. http://apps.leg.wa.gov/rcw/default.aspx?cite=69.51A.010.

[2] The Tor Browser Bundle. https://www.torproject.org/projects/torbrowser.html.

[3] WA Senate Bill 5073. http://apps.leg.wa.gov/documents/billdocs/2011-12/Pdf/Bills/Session%20Law%202011/5073-S2.SL.pdf.

[4] Hands off Washington Patients, 2011. http://cdc.coop/registry.

[5] ACLU of Washington. Medical marijuana patient records are private, court rules, 2007. http://bit.ly/lPODeY.

[6] Auckland Stuff.co.nz. Staff pry into files of celebrity patients, 2009. http://www.stuff.co.nz/auckland/local-news/130205.

[7] Charles Ornstein. Fawcett's cancer file breached, 2008. http://articles.latimes.com/2008/apr/03/local/me-farrah3.

[8] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, August 2004.

[9] C. Dwork. Differential Privacy. In V. Sassone and I. Wegener, editors, *ICALP*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.

[10] C. Dwork. Differential Privacy: A Survey of Results. In M. Agrawal, D. Du, Z. Duan, and A. Li, editors, *Theory and Applications of Models of Computation*, volume 4978 of *Lecture Notes in Computer Science*, pages 1–19. Springer Berlin / Heidelberg, 2008.

[11] EFF. Aol's data valdez violates users' privacy. https://www.eff.org/deeplinks/2006/08/aols-data-valdez-violates-users-privacy.

[12] C. Essig. Illinois makes millions selling personal information, 2010. http://www.thesouthern.com/news/article_0a5fd6a0-4b6b-11df-a353-001cc4c03286.html.

[13] J. Estus, P. Monies, and G. Off. State profits from residents' data, 2010. http://www.tulsaworld.com/news/article.aspx?subjectid=11&articleid=20100404_11_A1_Thesta994848.

[14] J. A. Halderman, S. Schoen, N. Heninger, W. Clarkson, W. Paul, J. Calandrino, A. Feldman, J. Appelbaum, and E. Felten. Lest we remember: Cold boot attacks on encryption keys. In P. Van Oorschot, editor, *Proceedings of the 17th USENIX Security Symposium*, pages 45–60. USENIX, July 2008.

[15] N. Li, T. Li, and S. Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *International Conference on Data Engineering*, 2007.

[16] Marijuana Policy Project. Grid: A comparison of key aspects of state medical marijuana laws, 2011. http://www.mpp.org/assets/pdfs/library/MMJGrid15StatesMarch2011.pdf.

[17] National Cancer Institute. Marijuana Use in Supportive Care for Cancer Patients, 2010. http://www.cancer.gov/cancertopics/factsheet/support/marijuana.

[18] C. Percival. Stronger key derivation via sequential memory-hard functions. http://www.tarsnap.com/scrypt.html.

[19] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10:557–570, October 2002.

[20] Z. Wilcox-O'Hearn. http://en.wikipedia.org/wiki/Zooko's_triangle, 2003.

[21] WLWT News 5. IRS Worker Admits Snooping In Celebrities' Files, 2008. http://www.wlwt.com/news/17015370/detail.html.