Cyber Passport

Preventing Massive Identity Theft

Cryptographic Solutions; Administrative Framework.

Gideon Samid
Department of Electrical Engineering and Computer Science
Case Western Reserve University, Cleveland, Ohio
BitMint, LLC
Gideon@BitMint.com

Abstract: Identity Theft is the fastest rising crime in the United States with about 7% of US adult population victimized annually. This frightening scope warrants a bold government intervention. Here is a detailed proposal. "Cyber Passport" addresses itself to the main threat: a breach of a merchant, bank, or government department resulting in theft of identities of millions of citizens, which for a long time live in fear of residual violations. The solution is based on two principles: (i) online transactions may require a randomized, readily replaceable, short lived code (cyber passport); (ii) the cyber passport will be comprised of a working code, and of an un-stored code which is known only to the issuing agency and to the individual recipient. When implemented these two principles will prevent a massive violation - the biggest plague today. The un-stored code cannot be stolen from any business database because it is not stored there. Hackers will still be able to steal everything but they will have to go retail, no more wholesale theft. Cyber-Passport is not a panacea, but it brings the threat down to size. The program will be optional for citizens, and voluntary for participating establishments facing the public. It will require some legislation, a non-trivial administration, and the use of modern cryptographic technology. Albeit, an organic growth implementation plan is presented herewith.

Keywords—identity theft, cyber security, cryptography.

I. INTRODUCTION

The fundamental reality that invites today' massive data theft is the situation whereby tens of thousands of public facing online establishments, (agencies), store the private financial and personal data of millions of citizens. Hackers need to find one such institution where the security is lax, and that is enough. With the data raided from a single source the hackers can inch up to the next target. The data in all these data storages is pretty much the same. And most of it is long living. Credit card companies extended the life span of their cards, so that the stolen card data is valid for many years. Of course stolen social security numbers, and stolen dates of birth are valid forever. So victims can never relax. Millions sign up with various monitoring agencies constantly on guard for emerging instances of violations. It's not just money anymore. Hackers sell personal data to people who wish to buy something unsavory online, and choose to do it via another identity. People's reputation is ruined for no fault of their own.

To counteract this situation two things are needed: deny the data thieves the opportunity to violate millions of citizens in one successful raid, and further deny them the ability to exploit their spoils for the long run. This implies that hacking will be restricted to retail data theft, and to short lived profit from such theft.

This countermeasure is all encapsulated in the Cyber Passport proposal

II. THE CYBER PASSPORT PROPOSAL

The underlying idea is to anchor identities on an off-line code, which is randomized (un-guessable), readily replaceable (quick recovery), and of two parts: one "un-stored" and the other a working code. The two parts serve as a foundation for a cryptographic protocol that is designed to (i) prevent wholesale compromise of identities, (ii) enable confidential communication that resists the Man-in-the-Middle (MiM) attack, and (iii) offers quick recovery and replacement for any compromised code. The two codes together are referred to as the Cyber Passport.

We describe ahead (i) the administration of cyber passport, and (ii) the cryptographic foundation of the same.

A. The Administration of Cyber Passport

Following the necessary legislation the government will establish a cyber passport administration which will (i) issue, maintain, and secure the cyber passport codes of the applying citizens, (ii) respond to 24/7 queries about current passports, and (iii) enforce the proper behavior of all participating institutions with public facing websites.

The detailed activities are listed ahead.

High Level Description: the government issues to applying citizens a personal, non-transferrable short code comprised, say, of 3 letters and 5 digits (all randomized), regarded as the "un-stored" code, and a second similar size code regarded as the "working code". The codes are passed off line after some formal verification of identity. The receiving citizens will use these two codes when they connect to any

website where they find the icon of "participating in the cyber passport program". Such websites, per their own volition, if they are from the private sector, will apply to the government to participate in the program. Participating websites will receive the working code of their registrants, and a verifying code for the un-stored crypto passport. They will then engage the connecting citizen in a cryptographic dialogue that would convince them that he or she are in possession of their respective working code (without ever transmitting that code itself). Based on this working code the website and the connecting citizen will establish a cryptographic secure channel. Through this channel the citizen will pass his or her "un-stored" cyber passport. The participating website will not have a copy of the un-stored code, but instead will use its stored verifying code to verify that the connecting citizen is in possession of the un-stored code. Once the test is OK, the website and the citizen communicate freely using the secure cryptographic protocol they established before.

Should an individual suspect that his or her cyber passport has been compromised, he or she will apply for an instant renewed passport, which the connected websites will readily find out about in their nominal query.

Since the un-stored code is not stored by any merchant, bank, or other databases, except at the issuing authority, then by securing this one single database the government will guarantee that no wholesale theft of cyber passport will ever take place. And any retail theft, will be short lived because the codes themselves are short lived, and readily replaceable upon suspicion of theft.

1) Issue, Maintain, and Secure Cyber Passports

A dedicated administration backed by proper legislation will be established. Using non-algorithmic random number generators, the administration will issue a fresh, previously unused, number for all applicants. For example, a standard size of 4 letters, and 5 digits (easily memorable code) will cover more than 45 billion numbers, which is more than enough. And also a number that has negligible chance to be guessed.

The numbers will not be pre-stored, but generated on demand, and only then stored. There will be two codes: the 'un-stored' one and the 'working' code. The administration will work out a procedure by which each applicant is identified via off-line means, and delivered these codes also via off line means. This can happen via regular mail, or via biometric identification in a government office or via a commissioned branch of a bank, or other institution. It can be issued via states' motor vehicle administration. The recipient will then key in both codes to his communicating devices.

The administration will maintain two secure databases for the two codes. The entire proposal hinges on the premise that these two databases of all the codes can be sufficiently secured. In other words: the point in this proposal is that today's vulnerability where people's sensitive data is kept in countless databases across the country, is remedied by a situation where all this data is kept in one database center, protected by our best security people. It resolves the dilemma of the weakest link that voids the value of the high security in the other links. If there is only one link, one database center, it can be kept secure by matching our best and brightest against the best and brightest of our adversaries. Today their best and brightest match themselves against our worst and dumbest in the weakest link.

The administration will use a proper cryptographic procedure -- the blind verification procedure -- that allows one holding a verification code (but not the code itself) to verify that a communicating partner does hold the respective code. The blind verifier code will be distributed by the administration to all the participating merchants, banks, and government departments dealing with the public. The recipient establishments will keep a database of this verification code but not the un-stored code itself. A surfer trying to connect will be prompted for his cyber passport code and that code will be verified by the blind verifier code (cryptographic details ahead). The net result is that no hacker could raid this merchant or department and harvest millions of personal cyber passport codes because no merchant or department will have a database of that code. And the cryptographic design is such that holding the verifier code does not allow one to use the proving procedure and pass as the code holder.

The working code, by contrast will be distributed to all the participating establishments for them to use in a cryptographic procedure that creates a secure communication channel with any individual who holds the same working code (and with no other). The working code will allow the establishment and the connecting individual to build a per-session secure communication channel in which the individual will pass his un-stored code, which the establishment will verify using the verifier code.

Should there be a successful raid on any cyber establishment -- the respective working codes will all be re-issued, and the gain to the raiders will be voided. Should an individual suspect that his code was compromised, he or she will apply for a fresh one.

The codes will be short lived by design, so that even if a code is stolen without detection it will have a short effective life.

2) Query Response

The administration for cyber passport will be ready 24/7 to respond to queries about the verifier code and the working code of any individual. All participating merchants will have secure communication channels with the administration to effect these queries and their responses.

A participating establishment, upon being accessed by an individual, will check its own database to see if this individual applied for a cyber passport. The establishment will not take the individual word for it. If no entry is found for that individual then the establishment will real time verify that this individual has not applied for his or her cyber passport. If it turns out that the individual by that name did apply -- then the current party is fraudster! And attempt will be made to round up the suspect.

If the verifier code does not verify the submission of the connecting individual then the establishment will query the administration to check if a fresh code was issued.

3) Participation Management

No individual will be compelled to participate, and no private establishment will be mandated to take part. Government departments for their consequential dealings with the public will participate by law. The idea being that as the program unfolds more and more individuals will opt to apply for their private cyber passport, to protect their identity, their bank account, their medical information etc. And as this happens, then banks and merchants will find it of a great disadvantage not to offer their customers this national protection, and will in turn apply to participate and abide by the rules of conduct that will come with it.

Any participating individual will have the right to opt out at any moment, and the same for any private sector establishment. Once out, no code will have to be submitted, but none of its protection will apply.

B. The Cryptographic Foundation of Cyber Passport

Cryptographically speaking we have two players: an establishment, E, (a government department, a merchant, a bank, a medical office, etc.), and an individual, I. Any establishment, E, maintains two databases one of verifier-code, v, and one for the working code, w, for all its registered individuals.

Some individual, I, connects to establishment, E, announcing its name or its registration code, and requests to do business with E. E will respond with a "non-repeat" dialogue with I. The dialogue will exchange data that is not a repeat of any previous dialogue with that individual. The dialogue will convince E that the party on the other side is in possession of its working code w, and also establish a secret shared key with which to encrypt their bilateral communication for this session only. Once this secure channel is established it will be used for I to pass to E, the un-stored code, u. Now, E has no possession of u, but only the possession of the corresponding verifier code, v. With v E will confirm that the individual across the line indeed is in possession of u. Once completed, E and I can use their per-session secure channel to do their business.

We shall now further elaborate on the mentioned procedures: proof of working code, w, using w to establish a secure communication channel, and the verification protocol, using v to verify possession of u.

1) Verification Of Working Code

The details of this procedure may be found in reference [Samid, 5/2016]. The concept is as follows: The establishment, E, selects a one time used random number, nonce, r, and sends r to the individual I. I and E both compute a number q which is a combination of the working code, w and r: q = q(w,r), where the function q can be predetermined or specified ad-hoc. There is no secrecy to this function, and of course r is exposed, only w, and hence q are secret.

Looking at the binary representation of q, I will parse q to t successive distinct substrings, according to a pre-established, or ad-hoc procedure, which is not secret. I will then use a non-algorithmic random number generator (NARNG) to generate a random permutation of these t strings. This permutation q_t will be sent over from I to E.

E, on its side, will do the same for q and break it down to the same t substrings. Upon receipt of q_t, E will check if indeed q_t is a permutation of q: $q_t = T(q)$. If it is, then E will conclude that I is in possession of w, and will also infer the transposition key, K_t, that transposed q to q_t. $q_t = T(q, K_t)$. This bilateral new secret K_t will be the basis of the secret communication channel between I and E.

A hacker, H, without the possession of q will not be able to infer q from q_t because there are t! permutations, and because the division of q into the t substrings depended on the value of q, so that looking at q_t it is not clear how to divide it to t substrings in the first place. And because unlike w, q is different for every session, owing to the nonce, r, the information gleaned from previous sessions will not help H to infer w.

Illustration: An establishment E has mailed an individual I a secret PIN: w=7854. I, at some point tries to connect with E. She identifies herself, and in response E sends I a nonce: r= 2973, with instructions to compute q as the absoslute difference between w and r: q=|r-w|. Both I and E then compute q to be q=|7854-2973| = 4881. Expressed in straight binary:

q = 4881 = 1001100010001. E and I agree on subdividing q to t substrings by reading q from left to right, and incrementing the size of each successive substring thereby assuring that no two strings will be identical. The last substring may be of a size larger than +1 relative the previous substring. This will happen if the last string is identical with a previous substring. So:

q = 1-00-110-0010-001

The result is t=5. There are 5!=120 permutation of q. I then uses a non-algorithmic random number generator to identify the numbers 1-5 in a randomized order, say: K_t = 3,1,5,2,4. Accordingly, q_t will be constructed by moving the 3rd substring in q to position 1 in q_t, moving substring 1 in q to position 2 in q_t, etc.:

q_t = 110-1-001-00-0010

I will then deliver q_t = 1101001000010 to E. E does not know K_t, but it knows q (and also q_t), so E will now evaluate q_t to verify that it is a strict transposition of q. To do that E will first try to match the largest substring (#4 in q) over q_t. There are two locations where 0010 fits over q_t. E will first try the first one: 1101**0010**00010 (the bold letters denote the overlady). Then E will try to fit substring #5 in q: q_t = 1101**001000**10. This fitting attempt fails because it requires two substrings each of size 1 bit, which is not the case with q. So E will check the other fit for substring #4 in q: q_t = 1101001000**10**. Then E will try fro fit substring #5: q_t = 1101**001**000**10**. Then E will try to fit substring #3 in q (110): q_t = **110**1**001**000**10**. Then E will try to fit substring #2 in q (00): q_t = **110**1**001000010**. And then substring #1 in q fits right in. E then concludes that q_t is a proper permutation of q. And furthermore, E now knows K_t = 3,1,5,2,4.

H, the hacker is not aware of q, only of q_t. He would not know how to subdivide q_t to substrings because the division

was performed based on q. So H will have to try all possible combinations of dividing q to t substrings. Since the chosen procedure for dividing q to substrings depends only slightly on the contents of q (and mostly upon its size, |q|), the hacker will have a good guess of t (to satisfy $0.5t(t+1)=|q_t|$). And will face t! options, each leading to a different working code, w. By selecting the identity of the nonce, r and the procedure q=q(r,w) the users determine the value of t (and the security parameter t!), which is the measure of security for the procedure that verifies w, and protects its identity.

2) *Establishing a Secure Communication Channel*

By communicating q_t to E, I has also communicated to it the randomized transposition key K_t, which was generated through white noise or through other non-algorithmic means. K_t will hence be a good per-session shared secret to be used for the secure communication channel between I and E.

I and E will agree on unit size, h, (bits per units), which will define a block size b=h*t bits, where t is the number of substrings to which q was divided. The non-algorithmic transposition key K_t will then be used to encrypt the conversation between I and E block by block. For better security, K_t will be used in a more elaborate protocol. A simple substitution cipher will prevent detection of partial order of the transposed result . The security of this solution is based on the size of the transposed list -- t. Since t is fully controlled by E and I, they can adjust it per the contents of their conversation.

The native security of transposition is super-exponential -- factorial. For a nominal t=50, the hacker will face 3.04140932 E+64 possible permutations. For t=100, the number of permutation rises to: 9.332621544 E+157.

The power of this method is that regardless of the size of the shared secret, w, the selected nonce, r, will determine the size of q, which in turn will determine the size of t. In practice this means that the two communicating parties will be in a position to decide the level of security they apply to every piece of communicated data. This is an important distinction compared to today's practice where the security is fixed by the used cipher, and is the same regardless of the sensitivity of the contents.

3) *Unstored code Verification*

Once E and I have established their secure communication line, they can use it for I to communicate the un-stored code, u, to E. E is not in possession of u, but is in possession of the u verification code, v=v(u). E will process the u value sent by I and compute its verification code, v'. If v=v' then E will conclude that I is in possession of the un-stored u value, despite the fact that E is not aware what it is.

III. Implementation Outlook

The matured version of the Cyber Passport program is not an immediate prospect. It requires legislation, notoriously a slow process, it requires a government administration -- another tedious proposition, and it requires the expected slow organic growth. It may be a full decade until that maturity is achieved. However, the nature of this proposal is such that it can start small, and through various independent implementation sites, and then grow organically, implementation-site by implementation-site, which will then fuse into a larger implementation.

An "implementation site" is an organization of a group of establishments E_g, offering to their combined customers, or registrants the option to apply for a cyber passport per that group (E_g). The E_g will invest in a joint cyber passport administration that would issue these codes, and be ready, 24/7, to verify them. Applicants (I) for the code will pay for the service since they are its beneficiaries -- greater security. That application fee may be complemented by an investment from the group of establishments, E_g, since the establishments also benefit from this program -- offering their customers and registrants a secure environment, which will serve as a powerful competitive edge versus others who don't offer the same. So between the application fees, and the E_g investment the cyber passport program may be clearly viable.

An implementation site may start humbly: a small number of establishments get together, with a small number of individuals (I) coming on board. But as this system operates, with almost zero burden, while more and more instances of would-be fraud cases are being prevented, because the fraudster did not show the right passport credentials, then through media outlet and dedicated advertisements, more and more establishments, on one hand, and more and more individuals, on the other hand, will apply to join the system. A positive inertia will develop. Then two or three developed implementation sites will join into a larger system, and so on.

The success of this grass root development will bubble up to the national initiative to implement the cyber passport program nation-wide.

Examples of Implementation sites: A group of online merchants, several banks, a pioneering state, medical establishments, a few federal departments. etc.

IV. Reference

Hochstein 2016 "Identity Theft - The Encyclopedia of Crime and Punishment" - Wiley Online Library

Jakobsson, 2007 "Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft" Indiana University Press.

Samid, 6/2016 "Time for a Cyber Passport" Digital Transactions Magazine June 2016

Samid, 5/2016 "T-Proof: Secure Communication via Non-Algorithmic Randomization" https://eprint.iacr.org/2016/474

Samid, 2009 "The Unending CyberWar" DGS Vitco http://www.amazon.com/Unending-Cyberwar-Gideon-Samid/dp/0963522043

Vikbladh 2016 "Identity Theft" Science 01 Apr 2016: Vol. 352, Issue 6281, pp. 46