Towards a user mitigated trade of privacy for web services

Edinah K. Gnang

April 1, 2013

1 Introduction

Privacy (broadly speaking) in the electronic age has become a valuable commodity as pointed out in [1]. Theoretical investigations of computational privacy games have been dominated by database centric models epitomized by the ϵ-differential privacy model introduced and developed by Dwork et all [5]. The inherent connections between privacy cryptography and mechanism design has also been emphasized in [2]. The database perspective on privacy focuses on providing anonimity garranties. In the present discussion we take up issue with some basic and often implicit assumption of the prevaling computational privacy paradigm.
- The individual user must entrust the database adminisatrator to implement users privacy policies.
- The goal of mechanism design for privacy should be to achieve a variant of "thruthfullness" inspired by mechanism designs for auctions.
- The variation amoung users privacy policy preferences are negligible.
- There exist no conflict of interest in the either of the participant roles that is either the beneficiary of the web services users and the providers of the services.
We suggest here that assuming that the web service market is set up in such a way as to promotes competition amoung the various service providers, it is easily possible for the users to conveniently implement and manage their privacy policies whithout entrusiting such policies to the service providers.

2 The Blakley-Shamir secret sharing scheme

The Blakley-Shamir secret sharing scheme was independently proposed by Blakley and Shamir in [1,3]. The present discussion is basic application their proposed sceret sharing scheme, but in order to make it self contained, we will review the basic notions. Both schemes fortunately can be described in terms of a system of linear equations of the general form

$$\mathbf{A}\,\mathbf{x} = \mathbf{b}. \tag{1}$$

We account for practical ressource limitations by assuming that the entries of $\mathbf{A}$ are elements of a sufficiently large finite field $\mathbb{F}_q$ and that the matrix $\mathbf{A} \in \mathbb{F}_q^{n\times n}$ is a randomly selected invertible matrix. For the purpose of the scheme we select a random vector $\mathbf{x} \in \mathbb{F}_q^n$. We may assume whithout loss of generality that the secret message we wish to share can be encoded as positive integer less than q and a single randomly chosen entry of the random vector $\mathbf{x}$ is substituted for the encoded message.
The splitting part of the scheme consists in providing each recepients a row of the matrix A as well as the corresponding row entry of the left hand side vector $\mathbf{b}$.

$$\{(\mathbf{a}_k,\, \langle \mathbf{a}_k\,, \mathbf{x}\rangle = b_k)\}_{0\leq k<n} \tag{2}$$

The message is therfore recovered by solving for the vector $\mathbf{x}$ and applying the decoding routine to the entries of $\mathbf{x}$. Since all but one of the entries of $\mathbf{x}$ are chosen at random, then with high probability there will be just one entry of the vector $\mathbf{x}$ whose decoding results in a valid message.

2.1 Analysis of the Blakley-Shamir secret sharing scheme.

The analysis of the Blakley-Shamir secret sharing scheme in our context is performed by upper-bounding the probability of two critical events. The first of which is the event that a single recepient guesses the remaining entries of the matrix in order

to implemenent a poly-time linear algebraic decoder. The single participant attack here uses the knowledge of the fact that the matrix is invertible to prune the search space and even so the probabilty of recovering the remaining entries of the matrix is bounded bellow by

$$\text{Prob}_1 < \left((q^{n-1}-1)(q^{n-1}-q)(q^{n-1}-q^2)\cdots(q^{n-1}-q^{n-2})\right)^{-1} \tag{3}$$

The bound follows from the fact that any $(n-1)\times(n-1)$ minor of A must be invertible and it is well known that the probability that an $(n-1)\times(n-1)$ matrix be invertible is precisely precribed by the right hand side of the inequality. The second event for which we want to provide a upper-bound of success is the the $n-1$ participant attack on the secret sharing scheme. Any $n-1$ participant attack will also use their knowledge of the fact that the matrix must be invertible in order to implement a linear solver decoder, the participant know that coefficient of the vector along the orthgonal complement to the vector space spanned by the rows which are in their possesion can not be zero and devise a strategy to guess the entries of the reamining vector. The $n-1$ attack succed with probability bounded by the following expression

$$\forall \epsilon > 0 \quad \text{Prob}_{n-1} < \left((q-1)q^{n-1}\right)^{-1} + \epsilon \tag{4}$$

As suggested by the analysis in each of these two extreme case the attack benefits very little from the partial information, in the sense that a linear algebraic attack based based on implementing a decoder using the partial information gathered by at most $n-1$ participants and uniformly guessing the entries of the last vector has less chance of success than discarding the partial info and simply trying to guess the answer. However it must be emphasized that while guessing the message has higher probability of success in this setting by discarding the partial information the attackers have no way of validating any particular guess.So the leverage of an $n-1$ attack is the ability to validate a guess.

3 Whistleblower and tattle-tale with set thresholds

We discuss here the first immediate application of the scheme. However we consider a very slight modification of the scheme to take into account predetermined participant number and time treshold. Alice whishes to share with her group of n friends a secret message which will be entrusted to her by her service provider. The service provider would entrust the secret with Alice under the condition that no less than the set fraction $\frac{n}{k}$ of the members of the group have access to the message within a predetermined set amount of time Δt of each other. Finally the set up should be such that any members' ability to read the message should constitute a certificate establishing that the message has also been read by the desired fraction $\frac{n}{k}$ of the members of the group within the predetermined amount of time Δt while not necessarily revealing the identity of the friends who also read the message.
On the message manager side at the first request he chooses a random $n\times n$ matrix of such that any selection of $\frac{n}{k}$ of the row has maximum rank while the whole matrix also has rank $\frac{n}{k}$ and at the same time he initiate the counter and provides that Alice and the requesting user each with the first and second row of the matrix respectively and the corresponing right hand side entry, for all subsequent request if they arrive withing the alloted time Δt each of the distinct requesters will be given a distinct row of the matrix $\mathbf{A}$ and the corresponding entry of the left hand side vector. After Δt Alice checks back with the verifier after a certain amount of time. If Alice can read the message than she can safely conclude that the desired fraction of members of the group have also been able to read the message within the prespecified amount of time.

4 "Pop-In" blockers

In the early days of the internet, our web browsing experience was sometimes inconvienienced by the spontaneous appearence of unsolicited Pop-up ads. Incidentally most web browser have incorporated what has come to be known as Pop-up blockers. In the current set up of electronic sevices which include electronic mail services, electronic social networking services, or distributed storage services, legitimate concerns pertaining to user privacy have been repeatedly raised. Many of these concerns can be summarized into two braod themes. The firts theme ecompasses the the service provider confilct of interest position. The conflict of interest here results from the fact that users entrust the service provider to enforce privacy policies which quite often conflicts with their advertisment revenue incentives. The second theme is concerned with the non explicit character of the ongoing trade of electronic services for private user information. In all faireness it should be appreciated that assesing the market value of user information while providing absolute grranties for user privacy is a daunting task for any service provider for reasons which might include
- The fact that electronic service providers do not always appriori know how to monetize the services they propose.
- Second there is a growing body of knowledge which suggest that anonimizing large database while retaining the relevant

information proves to be computationaly conflicting requirements.
- The users have varying range of privacy requirements so that one setting to be fitted to all may not be adequate for all the users.
- Service side focus policies for privacy unfairly puts the privacy burden on the providers while de-emphasizing the role of the user in the computational privacy game.
Havin pointed out these issues it appears of interest to identify specific privacy issues and suggest for them some computational of mechanism design solutions. A good illustration of the kind of privacy intrusion which have become quite routine and quite concerning are electronic scans of personal email correspondances by service providers like Google, Yahoo or Hotmail. Large corporation such as Google state on their website that their Ad targeting advertisement in their electronic mail services called Gmail is fully automated, and no humans read the emails.To make matters even worse relatively large education institutions such as Rutgers have plan to transfer the management of electronic mail services to companies to Google whithout really educating the community on the privacy policies put in place to protect the correspondance of their relatively large body of students. The real trouble with service provider implementation of privacy policies is the fact that almost invariably the user is left with virtually no means of assessing the service provider privacy practicing, for instance privacy audit program well established in the health record industry are virtually innexistant in the consumer electronics industry.
By analogy to Pop-ups, we call Pop-ins any form of unsolicited scans of personal electronic mail correspondance or any other form of personal electronic communication stream. Just as Pop-ups prompted the development of pop-up blockers so too we suggest here a simple procedure based on the Blakley-Shamir secret sharing scheme to be used to implement pop-in blockers. In our setting for simplicity we consider two individuals who whish to communicate via electronic mail, but whish to have a provable garantee under mild assumptions that the electronic service provider will not read their electronic correspondance. To achieve our goal we assume that each of the parties have $s > 1$ e-mail addresses from competing electronic mail providers. We remark that such a requirement is not at all unresonable since the e-mail service providers themselves tacitly assume that users have more than 1 email addresses as illustrated generic password recovery mechanisms. We claim that under these natural assumptions the user can use the Blakley-Shamir to send a messages that can be read only by sender and the receiver. It suffice for the sender to send s messages each containing $(\mathbf{a}_k, \langle \mathbf{a}_k, \mathbf{x} \rangle)$ where the $\mathbf{a}_k$ denote the k-th row of a randomly chosen invertible $s \times s$ matrix $\mathbf{A}$ and $\mathbf{x}$ denotes a random vector in which the message m is embeded as a field element. As discussed in previous sections A linear solver based decoder for the message for $s-1$ attack will succeed with probability bounded by $\left((q-1)q^{s-1}\right)^{-1} + \epsilon$ for all $\epsilon > 0$. Without loss of generality say that Bob sent to Alice a s-secure message. This approach is not significantly different from simply encrypting the messages and sending sending message and the key in two different email correspondances using different services, the main advantage of the Blakley-Shamir secret sharing scheme proposed here is convenience and it's versatility. For instance it enables us to send an t-secure message by sending with $t+r$ messages such that any combination t messages ensures that the reader can recover the message, in other words there is little symmetry breaking between the messages while in the encryption approach informally speaking there is a fundamental breaking of symmetry between the messages and the key. As a result message encryption and sending keys through different channel makes it more difficult to implement a combinatorial message recovery mechanism comparable to the Blakley-Shamir secret sharing scheme.

4.1 Sage-Python Implementation of the E-mail scan blocker

We now present for the sake of completeness a straight forward implementation of the Adi-shamir encryption scheme.
the workflow consists in converting ASCI characters in a text file into digits of a number. The **file2number**(filename) function takes as argument the name of a file corresponding to the E-mail which is assumed to be located in the same folder as the script defining the function. It turns the file into an integer. This is done by considering the text to be a large number expressed in base 256 thereby associating a digit with each one of the possible 256 ASCI characters. Many other ways can be used to map messages to integer, however the essential property of such mapping is that such mapping be bijective between integers and text messages.

```
def file2number(filename):
    n = 0
    f = open(filename, 'r')
    s = f.read()
    for i in range(len(s)):
        n = n + ord(s[i])*256^i
    f.close()
    f = open('out_'+filename, 'w')
```

```
    f.write(str(n)+'\n')
    f.close()
    return n
```

The **number2file**(n, filename) function inverts the bijective map prescribed by **file2number** and implemented above. It takes as argument an integer and the name of the desired name of the output file in which will be written the text associated with the decoded message. The implementation of the function **number2file**(n, filename) is provided bellow

```
def number2file(n,filename):
    s = ''
    i = 1
    while n>0:
        s = s+chr(n%256)
        n = (n-n%256)/256
        i = i+1
    print s
    f = open(filename, 'w')
    f.write(s)
    f.close()
```

The main function here is the **shamirize_trsh**(output_file, securty_level, treshold) function. The function takes three inputs respectively corresponding to the input file containing the original message, the security level for the secret sharing which corresponds to minimum number of points required to recover the message and the treshold input correspond to the number of files to be outputed files.
It is important to remark that the treshold input which builds into the scheme some redundancy must be greater or equal to the security level. The names of the files are chosen to be random integers. As to the coefficients of the polynomial function is selected at random by selecting the integer coefficient randomly except for the constant term which will correspond to the message.

```
def shamirize_trsh(input_file, scrt, trsh):
    # scrt denotes the security level for
    # the secret sharing scheme
    # trsh denotes the treshold of the number
    # of part necessary to recover the message
    var('x')
    m = int(file2number(input_file))
    fnct = int(str(m).replace('L',''))
    print fnct
    for i in range(1,scrt):
        v = int(str(randint(m+1,m^2)).replace('L',''))
        fnct = fnct + v*x^i

    # Writting the number into a the trsh files.
    for i in range(trsh):
        filename = str(randint(1, 10^5))
        s = filename+'.txt'
        xk = randint(1, 10^7)
        f_xk = fnct.substitute(x=xk)
        f = open(s, 'w')
        f.write(str(xk).replace('L',''))
        f.write('\n')
        f.write(str(f_xk).replace('L',''))
        f.write('\n')
        f.close()
```

Having implemented the functions which splits our secret message, we now turn our attention to functions which would enable us to recombine the message from the receivers end The function **interpolating**(List) takes as input a list of strings

which corresponds to the names of the files which containts data points to be interpolated. The function implemement the naive lagrange interpolation formula.

```
def interpolating(L):
    # L denotes the list of filenames
    X = []
    Y = []
    for i in range(len(L)):
        f=open(L[i],'r')
        tmp = ((f.readline()).replace('\n',''))
        X.append(int(tmp))
        tmp = ((f.readline()).replace('\n',''))
        Y.append(int(tmp))
        f.close()
    # Lagrange Interpolation part
    var('x')
    fnct = 0
    for i in range(len(L)):
        rg = range(len(L)); del rg[i]
        prd = 1
        for j in rg:
            prd = prd*(x-X[j])/(X[i]-X[j])
        fnct = fnct + Y[i]*prd

    # Returning the constant term
    return fnct.substitute(x=0)
```

The function **split_msg_trsh**(filename, security_level, total) is the main function called by the originator of the message, the three inputs of the function correspond to the file containing the message to partition, the security level which asserts the minimal number of datapoints required to recover the message while the last argument corresponds to the total number of files produced for redudancy.

```
def split_msg_trsh(filename, prt, trsh):
    shamirize_trsh(filename, prt, trsh)
```

The last function **recombine_msg**(List, filename) performs the task for the reader of recombining the message from the parts. Its inputs are the list of strings which correpsonds to the names of the files which contain the data point while the last input whish is a string correspond to the name of the file in which to ouput the recovered message.

```
def recombine_msg(L, filename):
    n = interpolating(L)
    number2file(n,filename)
```

5 Conlusion

While there are great variety in the in kind of electronic services available today we whish to argue that one abstract constant in most electronic service is the fact that users are produce content thereby comitting sensible private information which become valuabe commodities. (In the ongoing trade of private data to be mined the exchange for electronic services approach to privacy) it proves to be hard to accomodate all the privacy requirements while retinaing the wealth of valuable information. In addition there are great level of variability in the users privacy expectation. Electronic service providers, in regard to the privacy goals of their users are in obvious position of conflict of interest.
As illustrated by the relative sophistication of mechanism designs for add auctioning transaction that are carried out as a

result of users activity and while terms of the transaction between coroporation providing electronic goods and services and the ones buying add space and buying into the sophisticated costumer targeting system are explicit. It is becoming more and more appearent that the terms of the transactions between the receivers of the electronic goods and their providers are no so explicit. And why should these terms be clear, while the add commodity are clearly valued, there are virtualy no valuation markets accessible to the individual users for their information. To make matters worse, it has been histirocaly the case that the technolgy seems to be evolving faster than we can appreciate the value and the perils of the services being provided. Finaly the valuation of private information seem to depend on the ability for the broker to aggregate and organise the information. While we recognize that it might be difficult to propse absolute valuation for private informations we aregue that the aim of mechanis design should be to provide incentive to both the user and the service providers to make explicit the terms of the trade of private information for elctronic services.

We discuss here a strainght forward application of the Blakley-Shamir secret sharing scheme for Privacy/Service mechanism design destined to allow users of electronic services such as electronic mail(E-mail), social networking, forum post, to implement and enforce their own privacy goals. There is no doubt that more scheme relying on inisight in cryptography will be coopted in the future to propose more creative ways to restore the bargaing power of the user in the ongoing trade of private information for elctronic services.

In this paper the authors tried to make the case for user centric computational privacy game models which empower the user to implement their privacy goal. It is forseable that as user become more aware of computational aspects of privacy management and adopt more cautious privacy attitudes, this change might deny some electronic providers the acces to the ressources required to make the electronic service available to their users free of monetary charges. Therefore a compromise will have to be reached for the trading of electronic services for private information to continue. The goal therefore of proposing the meachanisms which allow the users to implement their privacy is to provide machanism which ultimatly lead to more explicit terms in the transaction between the service providers and their users.

Acknowledgments

The authors are grateful to Professor Doron Zeilberger, Professor Ahmed Elgammal Vukosi Marivate and Ryan Integlia for insightful discussions and suggestions. The author was partially supported by the National Science Foundation grant NSF-DGE-0549115.

References

[S+11] W. A. Stein et al., *Sage Mathematics Software (Version 4.7.2)*, The Sage Development Team, 2011, `http://www.sagemath.org`.

[1] Adi Shamir. How to share a secret. Commun. ACM 22, 11 (November 1979), pp.612–613

[2] Aaron Roth, Arpita Ghosh. Selling Privacy at Auction. EC11 Proceeding of the 12th ACM conference on Electronic Commerce, pp.199–209

[3] Blakley G. R. Safeguarding cryptographic keys. Proceedings of the National Computer Conference 48, 1979, pp.313–317

[4] Maskin, Eric S. Mechanism Design: How to Implement Social Goals. American Economic Review, 98(3) 567–76, 2008.

[5] C. Dwork and M. Naor, On the Difficulties of Disclosure Prevention.or The Case for Differential Privacy, Journal of Privacy and Confidentiality (2), 2010