Sistema per protegir la privadesa dels usuaris dels motors de cerca d'Internet

Rubén Salvador Gareta
Enginyeria en Informàtica

Jordi Castellà-Roca

Desembre 2011

Índex

- Introducció
  - Objectius
  - Justificació
  - Planificació del projecte
- Descripció de l'esquema
- Disseny de la implementació
- Implementació
- Marc d'avaluació
- Resultats
- Conclusions

Introducció

- Molts dels usuaris dels motors de cerca d'Internet realitzen cerques sobre els seus interessos, la seva persona, la seva feina, els seus problemes, etc.
- Aquesta informació queda emmagatzemada en l'historial de cerques del perfil de cada usuari dins del motor de cerca per facilitar que el cercador serveixi la informació més ràpidament.
- Aquesta informació conté dades personals i privades que permeten identificar unívocament la identitat dels usuaris perdent d'aquesta forma l'anonimat.
- A partir de l'any 2006 es va començar a prendre consciència sobre la necessitat de fer servir mecanismes de protecció de la privacitat dels usuaris dins dels motors de cerca d'Internet de forma externa i independent als propis cercadors.

Introducció: Objectius

- Implementar l'esquema de protecció de la privacitat dels usuaris dels motors de cerca d'Internet proposat per Alexandre Viejo i Jordi Castellà-Roca en la seva publicació "*Using social networks to disort users' profiles generated by web search engines*".
- Els usuaris del servei implementat han d'estar organitzats i connectats entre sí a través de les xarxes socials, i han de compartir entre ells la informació i els serveis sense dependència de cap node central o entitat tercera que supervisi la informació intercanviada.
- Les peticions de cerca dels usuaris han d'anar viatjant a través de la xarxa social seguint una sèrie d'algorismes de distribució, fins que algun d'ells enviï la consulta directament al motor de cerca.
- S'ha de garantir la correcta protecció del sistema en front d'usuaris il·legítims que pretenguin utilitzar l'anonimat que proporciona el servei per a realitzar consultes il·legals.

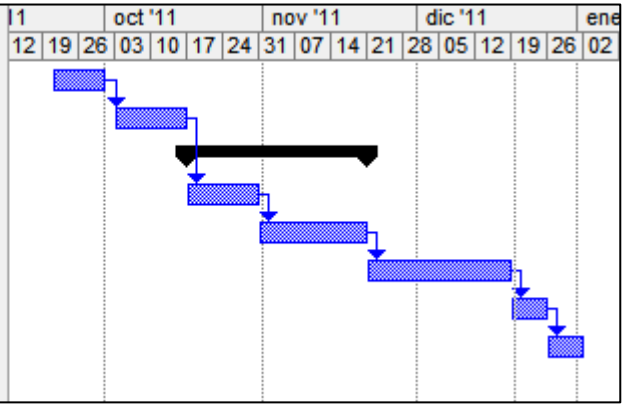
Introducció: Justificació

- L'esquema proposat per Alexandre Viejo i Jordi Castellà-Roca garanteix la privacitat dels usuaris dels motors de cerca.
- Es fonamenta en dos principis bàsics:
  - La proliferació de la connexió a Internet de banda ampla en tots els ordinadors.
  - La introducció gradual dels usuaris en les xarxes socials.
- Sistema independent d'entitats externes o espais de memòria compartida per a la distribució de les peticions de consulta.
- Els algorismes de distribució de la proposta eviten l'ús de consultes aleatòries o falses per a aconseguir l'emmascarament de les identitats.
- No són necessàries infraestructures complexes i poc usables.
- No és necessària col·laboració per part dels motors de cerca.

Introducció: Planificació

- El desenvolupament del projecte es va dividir en una sèrie de tasques. Al següent diagrama de Gantt es pot apreciar la planificació del projecte.

	Nombre de tarea	Duración	Comienzo	Fin
1	Redacció de la introducció	8 días?	mié 21/09/11	vie 30/09/11
2	Redacció de la descripció de l'esquema	10 días?	lun 03/10/11	dom 16/10/11
3	**Disseny i implementació**	**25 días?**	**lun 17/10/11**	**dom 20/11/11**
4	Disseny de l'esquema	10 días?	lun 17/10/11	dom 30/10/11
5	Implementació de l'esquema	15 días?	lun 31/10/11	dom 20/11/11
6	Resultats i Jocs de proves	20 días?	lun 21/11/11	dom 18/12/11
7	Lliurament final	5 días?	lun 19/12/11	dom 25/12/11
8	Presentació defensa del PFC	5 días?	lun 26/12/11	dom 01/01/12



Descripció de l'esquema

- Model del sistema
- Protocol d'enviament de consultes anònimes
  - Funció Ψ: Estimació del nivell d'exposició del perfil dels usuaris
  - Funció Υ: Avaluació del nivell d'egoisme dels usuaris
- Anonimat dels usuaris
  - Mecanisme de responsabilitat
- Adversaris del sistema

Descripció de l'esquema

- **Model del sistema**

  - **Usuaris**: Individus que envien consultes cap al motor de cerca d'Internet. Organitzats en xarxes socials.
  - **Motor de cerca**: Servidor al qual els usuaris envien les seves cerques.
  - Un usuari U_i genera una consulta q_i per a ser enviada cap el motor de cerca d'Internet. Amb aquesta consulta l'usuari U_i pot fer dues coses:
    - Enviar la consulta q_i directament cap al motor de cerca.
    - Enviar la consulta q_i cap a un dels seus contactes de la xarxa social. Per exemple U_j
  - Un usuari U_j que rep una consulta q_i de l'usuari U_i pot fer amb ella dues coses:
    - Enviar la consulta q_i directament cap al motor de cerca.
    - Enviar la consulta q_i cap a un dels seus contactes de la xarxa social. Per exemple U_k

Descripció de l'esquema

• Protocol d'enviament de consultes anònimes

- Un usuari *U* amb una llista de contactes de la xarxa social $\{N_1,...,N_k\}$ pot delegar l'enviament d'una consulta *q* en un d'ells (o en ell mateix) enlloc d'enviar-la directament al motor de cerca.
- **Funció Ψ**: Estimació del nivell d'exposició del perfil dels usuaris
  - Retorna com a resultat a quin usuari de la llista d'usuaris $\{U,N_1,...,N_k\}$ li ha de ser enviada la consulta *q* per que aquesta sigui enviada al motor de cerca. Per exemple: N_j
  - Si l'usuari triat per Ψ és el mateix *U*, llavors U envia la consulta *q* directament al motor de cerca.
- **Funció ϒ**: Avaluació del nivell d'egoisme dels usuaris
  - Determina si l'usuari N_j que ha rebut la petició de consulta *q* enviada per l'usuari *U* ha d'acceptar-la o no.
- Si N_j accepta la petició: N_j inicia el protocol amb els seus contactes.
- Si N_j rebutja la petició: U inicia el protocol amb la resta dels seus contactes.

Descripció de l'esquema

- **Anonimat dels usuaris**

  - El sistema garanteix l'anonimat entre usuaris que no estan directament relacionats entre sí en la xarxa social.
  - Aquest anonimat pot animar usuaris deshonestos a enviar consultes il·legals al motor de cerca fent ús del sistema.
  - **Mecanisme de responsabilitat**: El sistema fa ús de les signatures electròniques avançades i la *PKI* per a protegir els usuaris honestos.
    - Cada petició de consulta q que un usuari U_1 envia a un usuari U_2 ha d'anar acompanyada d'un certificat de transacció signat amb el certificat electrònic de U_1: $\{q \,||\, U_1 \,||\, U_2 \,||\, timeStamp\}_{SKU1}$
    - L'usuari U_2 que rep la petició de consulta q pot demostrar devant una autoritat legal que la consulta q ha estat enviada per l'usuari U_1 presentant el certificat de transacció signat per U_1.

Descripció de l'esquema

• Adversaris del sistema

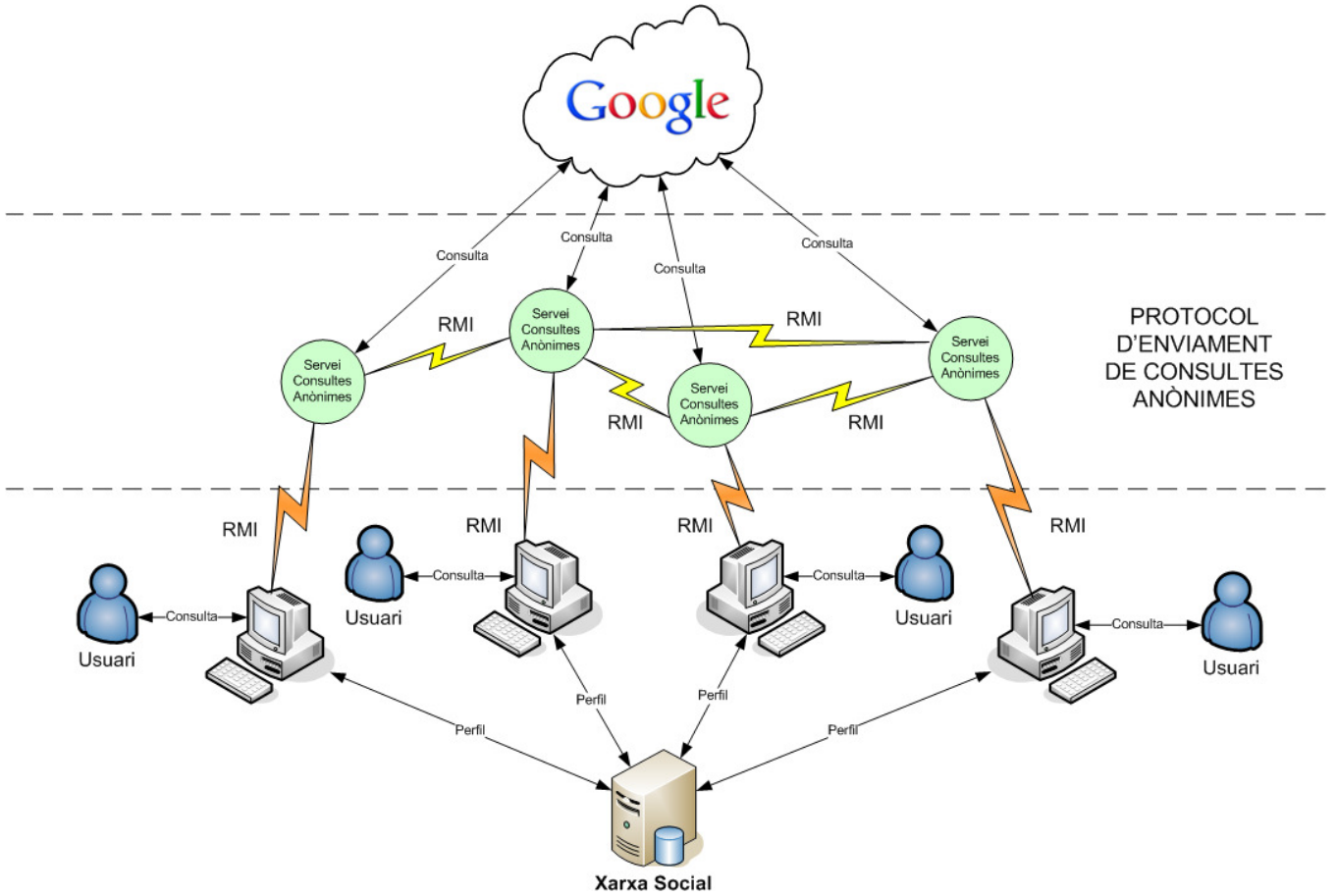
- **Motor de cerca**: El protocol d'enviament de consultes entre usuaris de la xarxa social evita que el motor de cerca pugui saber qui ha originat una consulta i per tant crear perfils dels usuaris.
- **Usuaris que creen perfils d'altres usuaris**: L'anonimat existent entre usuaris que no estan connectats a través de la xarxa social no permet que usuaris deshonestos creïn perfils de la resta d'usuaris a partir de les seves consultes.
- **Usuaris egoistes**: Els usuaris que mai accepten peticions d'altres usuaris són penalitzats per els algorismes de distribució del protocol d'enviament de consultes anònimes quedant aïllats i havent de enviar les seves pròpies consultes al motor de cerca.
- **Sybil attack**: Un usuari que crea usuaris falsos a la xarxa social no podrà entrar en el sistema a menys que la resta d'usuaris acceptin voluntàriament establir connexions amb ell.
- **Ususaris que envien consultes il·legals**: El mecanisme de responsabilitat del sistema evita que els usuaris enviïn consultes il·legals des del seu anonimat perquè aquestes han d'anar signades electrònicament.

Disseny de la implementació

- Esquema general
- Casos d'ús
- Infraestructura tecnològica
- Xarxa social
- Perfil d'usuari
- Protocol d'enviament de consultes anònimes
- Mecanisme de responsabilitat
- Comunicació entre usuaris

Disseny de la implementació

- **Esquema general**



Disseny de la implementació

Casos d'ús

- **Actors**:
  - Administrador de la xarxa social
  - Usuari del *Servei de Consultes Anònimes*

- **Casos d'us**:

  - Inicialitzar xarxa social: L'administrador de la xarxa social ha d'inicialitzar la xarxa social que el sistema farà servir posteriorment, amb les dades dels usuaris i les relacions existents amb aquests.

  - Arrencar el *Servei de Consultes Anònimes*: L'usuari del *Servei de Consultes Anònimes* que vulgui realitzar consultes anònimes contra el motor de cerca, en primer lloc haurà de posar en marxa el servei.

  - Enviar consulta al *Servei de Consultes Anònimes*: Una vegada el *Servei de Consultes Anònimes* ja està en marxa, l'usuari podrà fer consultes contra el servei enlloc de fer-les directament contra el motor de cerca.

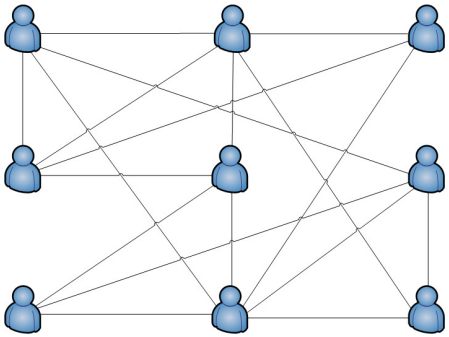
Disseny de la implementació

- **Infraestructura tecnològica**

  - Llenguatge de programació: *Java*
  - Llibreries addicionals
    - *Bouncy Castle Crypto API* (Operacions criptogràfiques)
    - *Apache Commons IO* (Operacions entrada/sortida fitxers)
    - *SLF4J i LOG4J* (Manteniment de *"logs"* del sistema)
    - *Oracle JDBC* (Connexió a base de dades)
  - Entorn de desenvolupament: *Eclipse Indigo IDE*
  - Eina de funcionalitats criptogràfiques: *OpenSSL*
  - Sistema de base de dades: *Oracle XE Universal*

Disseny de la implementació

• Xarxa social

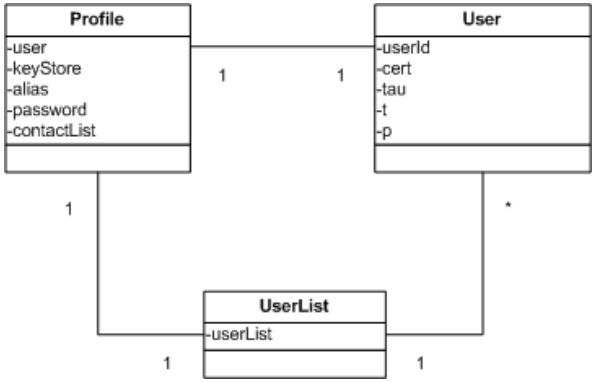


- Sistema de vinculació entre persones connectades entre sí per diferents tipus de relacions.
- Per a implementar l'esquema s'ha desenvolupat una simulació de xarxa social de la forma següent.
- **Implementació**:
  - Topologia de *xarxa en malla*. Cada node pot estar connectat amb la resta de nodes de la xarxa directament o indirecta.
  - Usuaris de la xarxa social identificats per la parella *<adreça IP:port>*
  - Cada usuari de la xarxa social té associat un certificat de clau pública.
  - Configuració de la xarxa a través d'un document *XML* que conté la informació dels usuaris i les relacions existents entre ells.
  - Dades dels usuaris emmagatzemades en una base de dades *Oracle Express* a la qual es connecta cada usuari per accedir a les seves dades.
  - Els usuaris es descarreguen de la xarxa social la informació del seu perfil d'usuari i dels perfils dels seus contactes directes.

Disseny de la implementació

• Perfil d'usuari

- Al iniciar el *Servei de Consultes Anònimes*, els usuaris descarreguen les dades dels seus contactes de la xarxa social:
  - Adreça IP i port
  - Certificats de clau publica
- Aquestes dades i les seves pròpies es carreguen a memòria per a permetre l'accés de forma ràpida al servei.
  - Adreça IP i port
  - Dades dels seus contactes
  - *Keystore* i dades d'accés al certificat i a la clau privada.
- Les dades descarregades de la xarxa social es guarden a la màquina de l'usuari en un document *XML* que s'anirà actualitzant.
  - `logs/profile.xml`




```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<User id="192.168.1.33:5552" p="1.0" t="0" tau="0">
      <Certificate>MIIDqzCCAxSgA...</Certificate>
      <ContactList>
            <User id="192.168.1.33:5551" p="1.0" t="0" tau="0">
                  <Certificate>MIIDqzCCAxSgA...</Certificate>
            </User>
            <User id="192.168.1.33:5553" p="1.0" t="0" tau="0">
                  <Certificate>MIIDqzCCAxSgA...</Certificate>
            </User>
            <User id="192.168.1.33:5552" p="1.0" t="0" tau="0">
                  <Certificate>MIIDqzCCAxSgA...</Certificate>
            </User>
            <User id="192.168.1.33:5554" p="1.0" t="0" tau="0">
                  <Certificate>MIIDqzCCAxSgA...</Certificate>
            </User>
      </ContactList>
</User>
```

Disseny de la implementació

• Protocol de consultes anònimes

- Cada node del document `profile.xml` descarregat per l'usuari correspon a un contacte. Aquest té associats uns atributs:
  - **p:** Probabilitat que té l'usuari d'acceptar consultes procedents del contacte corresponent al node.
  - **t:** Nombre de consultes que el contacte corresponent al node ha intentat delegar en l'usuari.
  - **tau:** Nombre de consultes que el contacte corresponent al node ha acceptat de l'usuari.
- Els valors d'aquests atributs seran carregats a memòria i utilitzats pels algorismes de distribució de consultes del servei:
  - **Funció Ψ**: Estimació del nivell d'exposició del perfil dels usuaris.
  - **Funció Υ**: Avaluació del nivell d'egoisme dels usuaris.
- Els valors d'aquests atributs s'aniran actualitzant i emmagatzemant sobre el propi fitxer `profile.xml`

Disseny de la implementació

• Mecanisme de responsabilitat

- S'ha fet servir *OpenSSL* per a generar els elements de la *PKI*: Certificats, claus i *"keystores"* de la Autoritat de Certificació i usuaris.
- Cada usuari disposa d'un *"keystore"* de tipus *PKCS#12* que es carrega a memòria durant la inicialització del *Servei de Consultes Anònimes*.
- El certificat de transacció $\{q \,||\, U_1 \,||\, U_2 \,||\, timeStamp\}$ s'implementa com un objecte *Java* serialitzat i codificat en *Base 64.*
- Les signatures del certificat de transacció generades són de tipus *CMS "attached".* Si és vàlida es comprova que les dades del certificat de transacció contingudes dins la pròpia signatura són vàlides.
- **Usuari envia petició a un contacte**: Genera un certificat de transacció i el signa amb la clau emmagatzemada al *"keystore"*.
- **Usuari rep petició d'un contacte**: Comprova que va acompanyada del certificat de transacció signat i valida la signatura amb el certificat de clau pública del contacte descarregat de la xarxa social.

Disseny de la implementació

• Comunicació entre usuaris

- Tecnologia de comunicació: *RMI (Remote Method Invocation)*
- Els sistemes dels usuaris fan de client i de servidor remot.
  - **Client:** Envia peticions de consulta al propi servei instal·lat a la màquina local de l'usuari i espera resposta.
  - **Servei:** Rep les peticions de consulta dels usuaris (del propi usuari o dels contactes directes) i s'inicia el protocol de consultes anònimes i el mecanisme de responsabilitat fins obtenir resposta.
- Al iniciar el *Servei de Consultes Anònimes*, els serveis remots dels usuaris es publiquen amb la adreça *IP* i el port de connexió especificats, quedant accessibles per a la resta d'usuaris.
- Els serveis dels usuaris localitzen els serveis dels seus contactes a través de les seves adreces *IP* i els seus ports descarregats de la xarxa social amb les dades del seu perfil.

Implementació

- Amb la compilació del codi del projecte mitjançant *Ant* s'obté el fitxer *ZIP* `anonymous-queries-1.0.zip` que conté:
  - `lib`: Directori amb les llibreries utilitzades per el *Servei de Consultes Anònimes* i la llibreria que conté la compilació del propi servei.
  - Comandes:
    - `StartSocialNetwork`: Comanda d'inicialització de la xarxa social.
    - `StartAnonymousQueriesService`: Comanda d'inicialització del *Servei de Consultes Anònimes*.
    - `StartAnonymousQueriesClient`: Comanda d'enviament de peticions de consulta al servei.
- Aquests comandes permeten executar els casos d'ús comentats amb anterioritat.
- Cada comanda rep una sèrie de paràmetres com els fitxer de propietats de connexió a la base de dades de la xarxa social, el fitxer de propietats d'accés al *"keystore"* de l'usuari, el port de connexió, etc.
- Tota la informació sobre el transcurs de la execució del *Servei de Consultes Anònimes*, consultes i certificats de transacció rebuts*, etc.* quedarà enregistrada als *"logs"* de la aplicació.

Marc d'avaluació

- **Objectius assolits**:
  - Analitzar en profunditat el problema de la conservació de la privacitat dels usuaris dins dels cercadors d'Internet.
  - Proporcionar una solució basada en la proposta d'Alexandre Viejo i Jordi Castellà-Roca en la seva publicació "*Using social networks to disort users' profiles generated by web search engines*".
  - Distribució de les consultes entre els usuaris de la xarxa sense necessitat de disposar de cap entitat tercera encarregada de supervisar ni distribuir la informació intercanviada entre els usuaris.
  - Garantir la correcta protecció del sistema en front a usuaris il·legítims o egoistes que pretenguin fer-lo servir.

- **Treballs futurs:**
  - Testejar la solució amb usuaris distribuïts per la xarxa.
  - Integració de la solució amb les xarxes socials existents actualment.
  - Integració de la solució amb els navegadors i sistemes dels clients.
  - Integració de la solució amb *PKI* reals.
  - Ampliació del mecanisme de responsabilitat (*OCSP*, *Smart cards*...)

Resultats

- **Privadesa:** Els algorismes de distribució implementats aconsegueixen que totes les consultes que els usuaris fan al motor de cerca quedin distribuïdes entre els seus contactes de la xarxa social, i per tant, el motor de cerca no pot saber quin usuari ha fet cada consulta i en conseqüència no pot construir el seu perfil. Quan la xarxa social creix en nombre d'usuaris i el nombre de contactes de cada usuari incrementa, el grau de privadesa també creix perquè, al haver-hi més usuaris entre els quals triar, les peticions queden més repartides entre tots ells i per tant, la seva identitat queda menys exposada.

- **Usabilitat:** Per parlar de usabilitat en termes de temps de resposta caldria executar el servei en diferents màquines distribuïdes per la xarxa. És un dels primers treballs futurs que caldria considerar.

- **Qualitat:** La resposta que el motor de cerca (en aquest cas *Google*) retorna al *Servei de Consultes Anònimes* quan aquest envia la consulta és gairebé la mateixa que la que retorna el motor de cerca directament, amb algunes petites diferències d'estil com són la absència de logos i d'alguns elements visuals. L'estil minimalista i amb escassos elements gràfics de *Google* ajuda molt en aquest aspecte.

Conclusions

- Amb l'ús del protocol *RMI* fent que els usuaris tant facin les tasques de client invocador del servei com de servidor remot, es va fer possible la correcta localització dels usuaris a través de la xarxa per part de la resta d'usuaris sense necessitat de disposar de cap entitat tercera.
- L'ús del format de signatura electrònica *CMS "attached"* per a la generació dels certificats de transacció signats va aportar complexitat al desenvolupament però va facilitar el transport de la informació en les transaccions entre usuaris i va estandarditzar la solució.
- Els requeriments d'emmagatzemament en que cada usuari del sistema ha de guardar els certificats de clau pública dels seus contactes i totes les consultes i certificats de transacció rebts van quedar coberts amb l'ús dels *"logs"* i del perfil d'usuari en format *XML* amb la informació de tots els contactes descarregat de la xarxa social.
- Encara quedaria treball per fer per a poder proporcionar una solució perfectament integrada amb els sistemes operatius i els navegadors dels usuaris, i amb les xarxes socials reals existents avui en dia.