A Peer-to-Peer Agent Community for Digital Oblivion in Online Social Networks

Klara Stokes and Niklas Carlsson
Linköping University, Sweden

Privacy, Security and Trust (PST 2013)
July 10-12 2013, Tarragona, Catalonia

Table of Contents

Table of Contents

1. Introduction
2. Design Goals
3. System Proposal
4. Considerations for Implementation
5. Protocols
6. Risks, Limitations and Future Work
7. Conclusions

Teenage Suicides Caused by Cyber-bullying

Amanda Todd killed herself at the age of 15 after repeated cyberbullying.

At age 13, she had contact with a 30- year-old man in a chat room who wan her trust and made her show her breasts for him.

Later, he contacted her on Facebook, and gave her an ultimatum: either she made a show for him or he would send the photo to everyone she knew.

She refused!

Teenage Suicides Caused by Cyber-bullying

Later the police knocked on her door telling her that the man had sent the photo to everyone she knew.

After this she was a victim for bullying in school and she tried to move.

But the man followed her over the Internet, and put up a Facebook page with her bare breasts as profile image, from which he contacted her new friends.

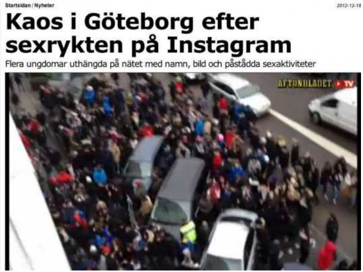
She was again a victim for bullying, physical and psycological, and again she changed school, but the cyberbullying was impossible to stop. Eventually she did not see any other solution than suicide.

Just before she ended her life, she tells her story in a Youtube video, holding up handwritten notes.

Teenage Suicides Caused by Cyber-bullying

Gothenburg, Sweden, December 2012

An account on Instagram containing explicit photos of teenage girls together with offending comments caused riots at a high school.

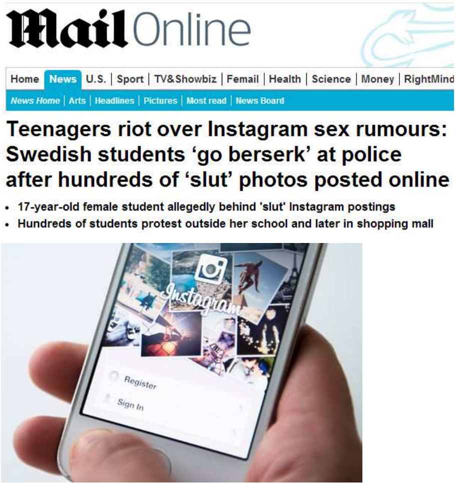


Startsidan / Nyheter

2012-12-18

Kaos i Göteborg efter sexrykten på Instagram

Flera ungdomar uthängda på nätet med namn, bild och påstådda sexaktiviteter



MailOnline

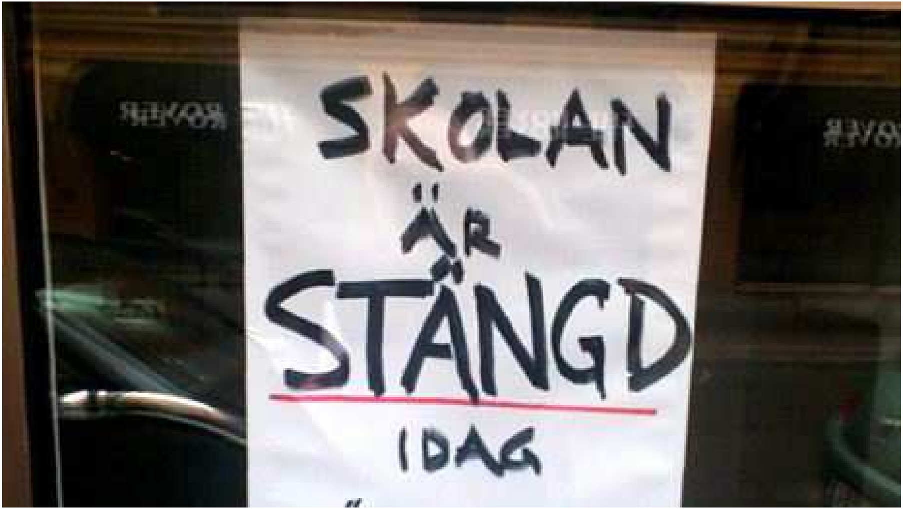
Home | News | U.S. | Sport | TV&Showbiz | Femail | Health | Science | Money | RightMind

News Home | Arts | Headlines | Pictures | Most read | News Board

Teenagers riot over Instagram sex rumours: Swedish students 'go berserk' at police after hundreds of 'slut' photos posted online

- 17-year-old female student allegedly behind 'slut' Instagram postings
- Hundreds of students protest outside her school and later in shopping mall

Gothenburg, Sweden, December 2012



"School is closed today"

Gothenburg, Sweden, December 2012

What now?

Young people want to use social networks.

In the past bullying took place at the school yard.

Today bullying is prospering on social networks.

When the school opens again, can it ensure that students can be on social networks without risk of bullying?

No! The Gothenburg Instagram photos were removed several times, but reappeared on new accounts.

These are not isolated events!

Teenagers upload or send explicit photos of themselves to others.

This is known as **sexting**.

Surveys show that 17.3% female teenagers has sent and 30.9% received such photos.

Male teenagers 18.3% sent and 49.7% received.

Why this difference?

These are not isolated events!

Teenagers upload or send explicit photos of themselves to others.

This is known as **sexting**.

Surveys show that 17.3% female teenagers has sent and 30.9% received such photos.

Male teenagers 18.3% sent and 49.7% received.

Why this difference?

Forwarding!

Removal of Offending Content in Facebook

On public demand, Facebook today supports the removal of offending content.

- Only content that explicitly violates the Facebook terms will be removed.
- An OSN administrator has to manually evaluate every removal request. Very time consuming!
- Typically, the OSN will try to remove the content within 72 hours.
- If the content does not violate the terms of the OSN, the user's only choice may be to directly ask the uploader to remove it.

Facebook also offers a "Social Reporting" tool, which allows the user to share the content that makes her uncomfortable with someone she trusts; e.g., a parent or a teacher.

However it seems these solutions still come short in practice...

The Right to be Forgotten

The right to be forgotten has been a hot debate subject for some years now.

Current opinion seems to suggest that feasible solutions for the right to be forgotten for today's Internet should use legal measures.

The European Commission has proposed (January 2012) a regulation intended to provide a right to be forgotten, still to be approved by the European Parliament (July 2013).

Digitial Oblivion

Let **digital oblivion** denote technical solutions for the right to be forgotten.

Current solutions often focus on attaching an **expiration date** on the published material.

Advantage: no need for the user to actively search for material that contain their personal information and that they might want to forget.

Digital rights management (DRM) has also been proposed for digital oblivion.

Our contribution: In contrast to expiry-based solutions, we take a pro-active approach and allow users to forget material within a restricted friendly domain. We assume the user previously found data through e.g.

- casual surfing,
- notifications by a friends, or
- by being tagged.

Table of Contents

1 Introduction

2 Design Goals

3 System Proposal

4 Considerations for Implementation

5 Protocols

6 Risks, Limitations and Future Work

7 Conclusions

Another Motivating Example

U and V both attend an event and W takes a photo of U and V together. Then W uploads this photo to an OSN, without the permission of U.

1. Assume that U wants to forget the photo, but either V , or W, or both, disagree and insist on that the photo should stay public. Who should decide?

Another Motivating Example

U and V both attend an event and W takes a photo of U and V together. Then W uploads this photo to an OSN, without the permission of U.

1. Assume that U wants to forget the photo, but either V , or W, or both, disagree and insist on that the photo should stay public. Who should decide?
2. Now assume that we agreed on who should decide, in terms of relation to the content. Now U claims that she has relation to photo that gives her right to remove it. How can this be verified?

Design Goals: Scenarios

- **Scenario 1.** The user wants to forget material she originally uploaded, appearing on her own timeline.
- **Scenario 2.** The user wants to forget material she originally uploaded, now appearing on someone elses timeline.
- **Scenario 3.** The user wants to forget material in which she appears, but which was not originally uploaded by her.

User-to-Content Relations

We claim that user-to-content (U2C) relations are critical for the correct design of a system providing digital oblivion.

The designers should answer at least the following two questions:

- Which U2C relations should give the user the right to decide that the content should be forgotten?
- How can these U2C relations be verified in a secure and automatic way?

Table of Contents

1 Introduction

2 Design Goals

3 System Proposal

4 Considerations for Implementation

5 Protocols

6 Risks, Limitations and Future Work

7 Conclusions

System Proposal

Let some users of a social network install a software agent with the following properties:

1. **Communication:** The agents of distinct users can communicate over a P2P overlay network.
2. **Filtering:** The agent is capable of
    - intercepting and modifying the material that the user uploads to the OSN, and
    - deciding what the OSN client will show to the user.
3. **U2C authentication:** The agent community is capable of establishing a protocol that allows for the authentication of some U2C relation.

Then the users that installed the agent can obtain a functionality of **digital oblivion within the community** they form and **with respect to the U2C relation** in question.

System Proposal

- The P2P community of agents creates a **virtual environment within the OSN** that will allow the users to claim digital oblivion of already published content.
- The virtual environment works as a **filter,** removing “forgotten” content from the OSN as it is observed by the users within the community.
- Users that do not install the agent will still be able to see the “forgotten” material.

Scenarios of Application

- When the users of the OSN quickly wants to remove annoying content with personal information, and the content is located on some other user's timeline. This may be a very attractive feature for many users.

Scenarios of Application

- When the users of the OSN quickly wants to remove annoying content with personal information, and the content is located on some other user's timeline. This may be a very attractive feature for many users.
- By making the use of the agent visible to the OSN friends of the user, usage of the digital oblivion functionality could be an ethical statement, helping the user to build up a positive online personality.

Scenarios of Application

- When the users of the OSN quickly wants to remove annoying content with personal information, and the content is located on some other user's timeline. This may be a very attractive feature for many users.
- By making the use of the agent visible to the OSN friends of the user, usage of the digital oblivion functionality could be an ethical statement, helping the user to build up a positive online personality.
- The functionality could be installed within the OSN. This would give added value to the OSN in question, in terms of user satisfaction.

Scenarios of Application

- When the users of the OSN quickly wants to remove annoying content with personal information, and the content is located on some other user's timeline. This may be a very attractive feature for many users.
- By making the use of the agent visible to the OSN friends of the user, usage of the digital oblivion functionality could be an ethical statement, helping the user to build up a positive online personality.
- The functionality could be installed within the OSN. This would give added value to the OSN in question, in terms of user satisfaction.
- Organizations often confront serious cyberbullying problems, in particular schools. They could include the digital oblivion functionality in their bullying prevention program. Facebook features Group for Schools, digital oblivion could be a requirement for joining the school online community.

Table of Contents

1. Introduction
2. Design Goals
3. System Proposal
4. Considerations for Implementation
5. Protocols
6. Risks, Limitations and Future Work
7. Conclusions

Considerations for Implementation

1. **Communication:** The P2P community can be built upon the topology of the existing network structure of the OSN, so that the agents of two friends in the OSN are neighbors in the P2P network.
2. **Filtering:** Possible solutions span from purely server-side (incorporating within the OSN the interception, modification and filtering of content) or client-side (application wrapper or browser plugin).
3. **Distributed U2C authentication:** Systems that allow distributed authentication of user-to-content relations, sometimes supported by trust mechanisms.

From a cryptographic perspective, perhaps the most interesting required property in our system is the third one: **Distributed U2C authentication.**

User-to-Content Relations

The following user-to-content (U2C) relations encode the motivating scenarios we saw in the beginning:

- **U2C R1.** User uploaded content to the OSN (Scenario 1 and 2).
- **U2C R2.** There is personal information on the user present in the content (Scenario 3).

Remember:

- **Scenario 1.** The user wants to forget material she originally uploaded, appearing on her own timeline.
- **Scenario 2.** The user wants to forget material she originally uploaded, now appearing on someone elses timeline.
- **Scenario 3.** The user wants to forget material in which she appears, but which was not originally uploaded by her.

Distributed Authentication of U2C Relations (U2C R1)

U2C R1. The user uploaded the content to the OSN.

U2C authentication through a combination of perceptual hash, digital signature and watermarking techniques, which allows to

1. embed a signed hash of the image into the image,
2. retrieve the hash in two ways from the image and compare these:
    - From the watermark, employing the user's public key.
    - Directly from the image.

Distributed Authentication of U2C Relations (U2C R2)

U2C R2. There is personal information on the user present in the content.

- **Tags** can indicate presence of personal information in OSN images. According to Facebook:

  "A tag is a special kind of link. When you tag someone, you create a link to their timeline."

  Main motivation for using tags in U2C R2 authentication is: indication comes from a user who has no interest, or negative interest, in the U2C R2 authentication.
- **Faces** also indicate presence of personal information in images. **Facial recognition** can be used to find specific faces in images automatically.
- For textual content, probably the best way to find indications of personal information is to use **semantics**, and other tools from **natural language processing**.

Distributed Authentication of U2C Relations (U2C R2)

Alone, these indicators of presence of personal information in content are too weak to provide secure U2C R2 authentication.

- Tags can be added with the intention to falsely indicate personal information where it is not present.
- Facial recognition can indicate presence of individual face in image, but information typically requires confirmation feedback from humans.
- Same is true for semantics in textual content.

Security can be added by using **trust management** in the agent community.

Table of Contents

1. Introduction
2. Design Goals
3. System Proposal
4. Considerations for Implementation
5. **Protocols**
6. Risks, Limitations and Future Work
7. Conclusions

1. **System start-up.** Assign each user identifier and public/private key pair.
2. **Upload of content.** Agent embeds watermark in the content.
3. **Request for oblivion of content.** Agent sends message to neighbors in limited broadcasting.
4. **Receiving oblivion request.** Agent finds content, verifies U2C authentication and indexes content on oblivion list.
5. **Oblivion viewing.** The agent filters away content that is stored on the oblivion list.
6. **Add a new friend.** Agent sends the user's current requests for digital oblivion to agents of new friends.
7. **Disapproval of oblivion of content.** Agent broadcasts disapproval of oblivion to community.
8. **Receiving disapproval of oblivion.** Agent evaluates reputation of disapproving agent and update trust values of involved agents.

High-level System Overview

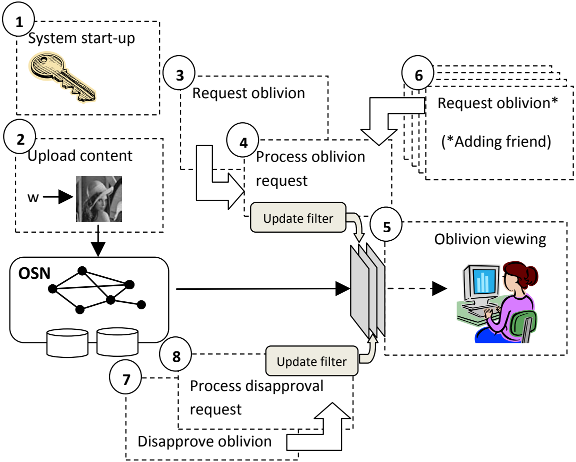


Table of Contents

Risks, Limitations and Future Work

Our implementation of digital oblivion is based on the distributed storage of lists that indexes content that have been requested to be forgotten.

A curious user within the community, or some malicious software that forwards oblivion lists to an adversary, could use it to identify embarrassing/hurtful content referring to the users in the community.

Future work should show how this risk could be mitigated through secure implementation, e.g. encryption of stored data.

Risks, Limitations and Future Work

Distributed digital oblivion requires that the user tells the distributed system what data she wants to forget.

Other users can find out what data she wants to forget.

A possible solution is to develop systems for anonymous U2C authentication.

Risks, Limitations and Future Work

Our system does not remove the data on the oblivion list from the real OSN.

An eavesdropper within the community could compare the real OSN and the oblivion view OSN, and localize content that should be forgotten through the differences.

We must rely on the good intentions of the participants.

In other words, the user is not friend with her harassers.

Our solution is meant to protect this friendly environment from outsiders.

Conclusions

- We give users of OSN
  - a **restricted functionality of digital oblivion**,
  - through a **distributed, user-managed** system
  - for **access control of content**,
  - based on **authenticable user-to-content relations.**
- Our system offers
  - **speed and autonomy**: users can implement the system without collaboration from the OSN,
  - **access control** also over content **on other users timelines**.
- The system is based on an agent community that together decide what content should be forgotten and that filter the user's view of the OSN.
- We have outlined a **candidate design** as a proof of concept.

Conclusions

Parts of our solution can be used separately, in other systems.

- A P2P agent community as a platform for collaborative security and privacy solutions.
- Combination of **perceptual hashes, digital signatures** and **watermarking** for use in digital oblivion.
  Applicable when the user wants to reclaim images that she originally uploaded, and others later uploaded again.
- Alternative methods for content uploaded by others. Combination of **indicator of presence of personal information** and **trust management.**
  Indicators of presence of personal information
  - in images: tags and facial recognition,
  - in text: semantics and language processing.

A Peer-to-Peer Agent Community for Digital Oblivion in Online Social Networks

Klara Stokes and Niklas Carlsson
Linköping University, Sweden

Privacy, Security and Trust (PST 2013)
July 10-12 2013, Tarragona, Catalonia