Introduction to Privacy Technologies

Claudia Diaz
KU Leuven – COSIC

Summer School on real-world crypto and privacy
June 2017

Overview

- What is privacy? (non-technical definitions)

- What are the “privacy concerns” in the context of technology?
- Which technical solutions exist to tackle those concerns?
- Challenges and limitations of those solutions

(Some) Definitions of Privacy

What is privacy?

- Abstract and subjective concept

- Dependent on:
  - Study discipline
  - Stakeholder
  - Social norms and expectations
  - Context

Warren & Brandeis (1890)

- From a legal perspective

- “The right to be let alone”
  - This citation was a response to technological developments (photography, and its use by the press)
  - Warren and Brandeis declared that information which was previously hidden and private could now be "shouted from the rooftops”

Westin (1970)

- “The right of the individual to decide what information about himself should be communicated to others and under what circumstances”

- “Informational self-determination” (German constitutional ruling, 1983)
  - “[...] in the context of modern data processing, **the protection of the individual against unlimited collection, storage, use and disclosure of his/her personal data** is encompassed by the general personal rights of the German Constitution. This basic right warrants in this respect the **capacity of the individual to determine in principle the disclosure and use of his/her personal data.**”

Agre and Rotenberg (1998)

- From a social psychology perspective
- “The freedom from unreasonable constraints on the construction of one's own identity”
  - The construction of one's identity is always mediated by “gaze of the other”
  - Impression management, self-presentation
    - Construct an image of ourselves to claim personal identity
- Social networks, profiling, search results

Solove’s taxonomy of privacy (2006)

- Information **Collection**
  - Surveillance
  - Interrogation
- Information **Processing**
  - Aggregation
  - Identification
  - Insecurity
  - Secondary Use
  - Exclusion
- Information **Dissemination**
  - Breach of Confidentiality
  - Disclosure
  - Exposure
  - Increased Accessibility
  - Blackmail
  - Appropriation
  - Distortion
- **Invasion**
  - Intrusion
  - Decisional Interference

Nissembaum (2004)

- From a moral philosophy perspective

- Concept of privacy as "**contextual integrity**"
  - The protection for privacy is tied to norms of *specific* contexts.

- Contextual integrity is maintained when both these types of norms are upheld:
  - Norms of **appropriateness**: what information about persons is appropriate to reveal in a particular context
  - Norms of **flow** or **distribution**: what can be done with that information (e.g., expectation of confidentiality)

- These norms may be
  - Explicit and specific
  - Implicit, variable, and incomplete

- Application to the evaluation of technical systems

Data Protection

- EU Data Protection Directive (1995)
- Data Protection Regulation (2016) will be in effect from May 2018

- Applies to "Personal data": any information relating to an individual
  - Does not apply to national security activities or law enforcement

- "Regulation on the **protection** of natural persons with regard to the processing of **personal data *and*** on the **free movement of such data**"

- Principles:
  - Transparency
    - Informed consent of the data subject, access rights
    - Necessity based on contractual, compliance, public interest, etc.
  - Legitimate purpose:
    - Personal data can *only* be processed for specified explicit and legitimate purposes, purpose limitation
  - Proportionality
    - Data must be adequate, relevant and not excessive in relation to the purposes for which they are collected and processed (aka "data minimization")
  - Accountability of the data controller

ECHR Art 8

- Emerged as a response to the excesses of totalitarian states in the 30s and 40s (entered into force in 1953)
  - Spirit: protect **citizens** from an overbearing/intrusive **state**
  - During the cold war: ‘western’ states would distinguish themselves from the ‘eastern block’ in that the population was not subject to pervasive surveillance

- European Convention on Human Rights Article 8 – *Right to respect for private and family life*
  - 1. **Everyone** has the **right** to respect for his private and family life, his ***home*** and his ***correspondence***.
  - 2. There shall be no interference by a public authority with the exercise of this right **except**
    - such as is in accordance with the law and is necessary in a democratic society in the interests of ***national security***, ***public safety*** or the ***economic well-being of the country***, for the ***prevention of disorder or crime***, for *the* ***protection of health or morals***, or for the protection of the rights and freedoms of others.

Related concepts

- Intertwined with other concepts
  - Freedom: anonymous speech, freedom of association
  - Dignity: airport scanners
  - Autonomy: censorship, filter bubble
  - (Non-)discrimination: profiling and personalization
  - Personal safety: identity theft
  - Democracy: targeted political messaging exploiting psychological biases

Privacy and Technology

Offline world → Online world

- Information is hard/costly to collect, store, search, and access
  - Conversation face-to-face
  - Letters in the post
  - Papers in an physical archive
  - Paying with cash
  - Following your movements
  - Knowing who your friends are
  - Looking for info in encyclopedia
- Information hard to copy/disseminate, easy to destroy
- Hard to aggregate, make profiles and inferences
- Information forgotten after some time
- …

- Information is easy/cheap to collect, store search, and process
  - Skype, instant messaging
  - Emails
  - Files in digital archive
  - Paying with credit card
  - Location tracking
  - “Online” friends
  - Searching in google, wikipedia
- Easy to copy/disseminate, but hard to destroy
- Easy to aggregate, make profiles and inferences: unique identifiers
- Information never forgotten
- …

Nothing to hide?

- Solove: “The problem with the ‘nothing to hide’ argument is its underlying assumption that privacy is about hiding bad things.”
- “Part of what makes a society a good place in which to live is the **extent to which it allows people freedom from the intrusiveness of others. A society without privacy protection would be suffocation.**”
- Difference between “secret” and “private”
  - Your daily routine, your movements, who your friends are, what you said in a conversation, which books you read…
  - These may not be secret, but you may not be comfortable with making it public or having external entities knowing about it, analyzing it, and extracting conclusions from it

Privacy and technology

- Bottom line: our actions and interactions are increasingly mediated by technology
  - We leave digital traces everywhere
  - Traces are combined, aggregated, and analyzed to infer further information about ourselves and to make decisions that affect us
  - We have no control over our information, or the inferences derived from it (lack of transparency)
- Information is never forgotten
  - But will perhaps be used out of context

Privacy Technologies

- Aim to address / mitigate certain privacy concerns
  - While allowing us to enjoy the benefits of modern ICTs

- Three categories of technologies and discuss:
  - Privacy concerns that motivate the solutions
  - Goals of the solutions
  - Example technologies
  - Challenges and limitations

“Social privacy”: Privacy concerns

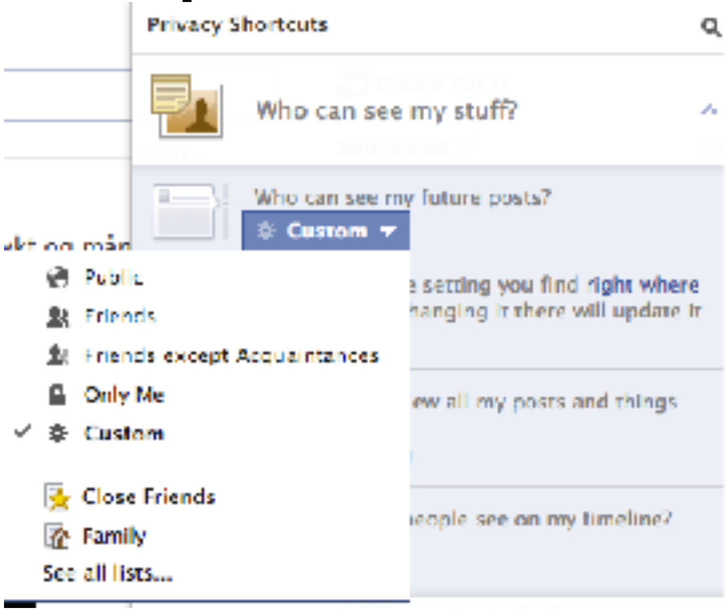
- Technology mediation of social interactions leads to problems in the immediate social context of the user
  - “My parents discovered I’m gay”
  - “My boss found out that I hate him”
  - “My friends saw my naked pictures OMG!”
- Self-presentation and identity construction towards friends, family, colleagues
  - Particularly relevant in social media applications
  - Tension between privacy and publicity
- Decision making: cognitive overload, bounded rationality, immediate gratification, hyperbolic discounting, behavioral biases
- **Who** defines the privacy problem:
  - Users

“Social privacy”: Goals

- Meet **privacy expectations**: *“don’t surprise the user!”*
- Make **privacy controls** (e.g., settings) visible and easy to use
- Support users in privacy-relevant **decision making**:
  - users can better predict the outcomes of their actions, such that they do not **regret** their actions after the fact
- Help users develop appropriate **privacy practices**
  - e.g., etiquette: use “Bcc:” instead of “Cc:” when sending email to a large number of people

“Social privacy”: Examples

- Appropriate defaults
  - “only friends”
- Usability of privacy settings
  - automated grouping of friends
- Contextual feedback mechanisms
  - “how others see my profile”
- *Privacy nudges*



Privacy Shortcuts

Who can see my stuff?

Who can see my future posts?

Custom

Public

Friends

Friends except Acquaintances

Only Me

✓ Custom

Close Friends

Family

See all lists...

e setting you find right where
hanging it there will update it

ew all my posts and things

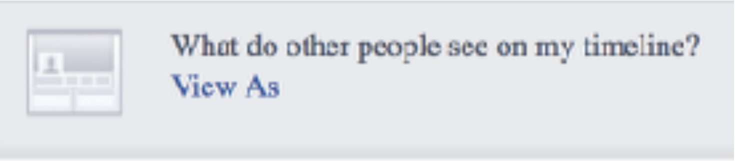
eople see on my timeline?

Post on your Timeline

Update Status Add Photos/Video

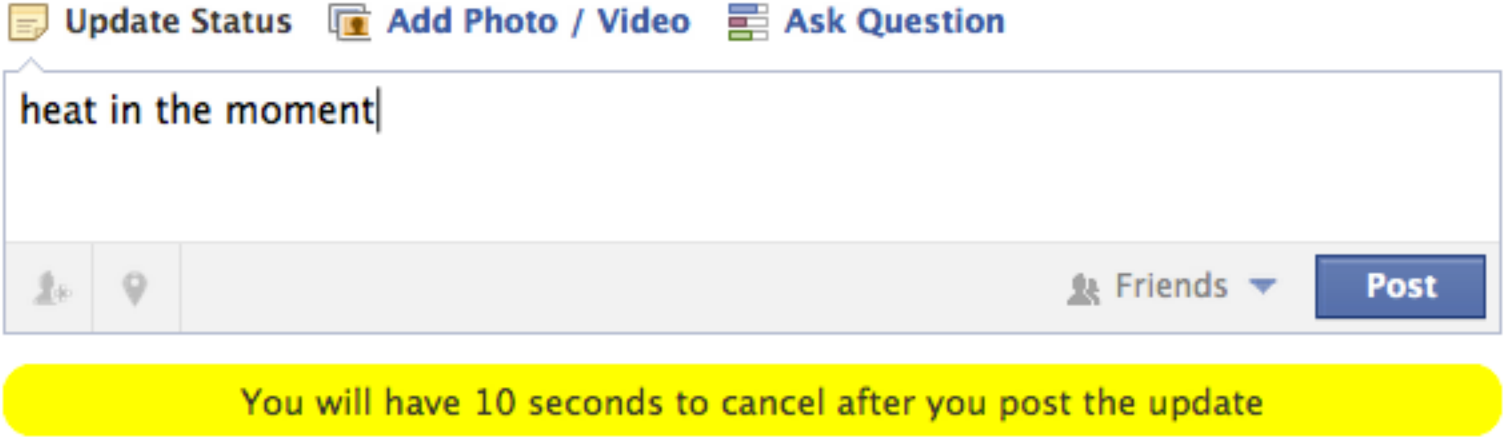
What's on your mind?

Post



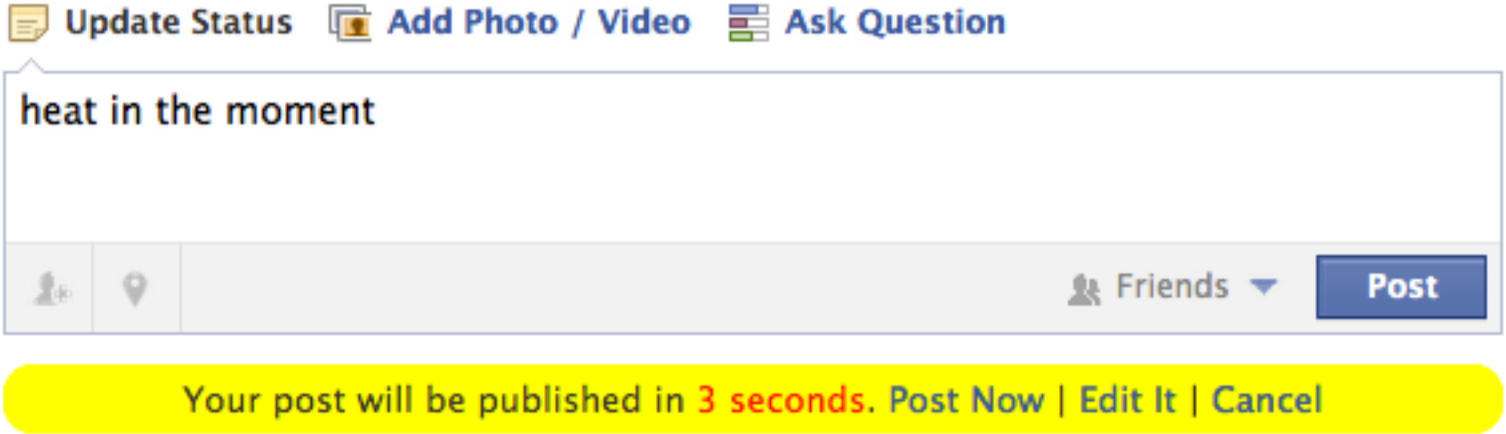
What do other people see on my timeline?
View As

Timer nudge (stop and think)



heat in the moment

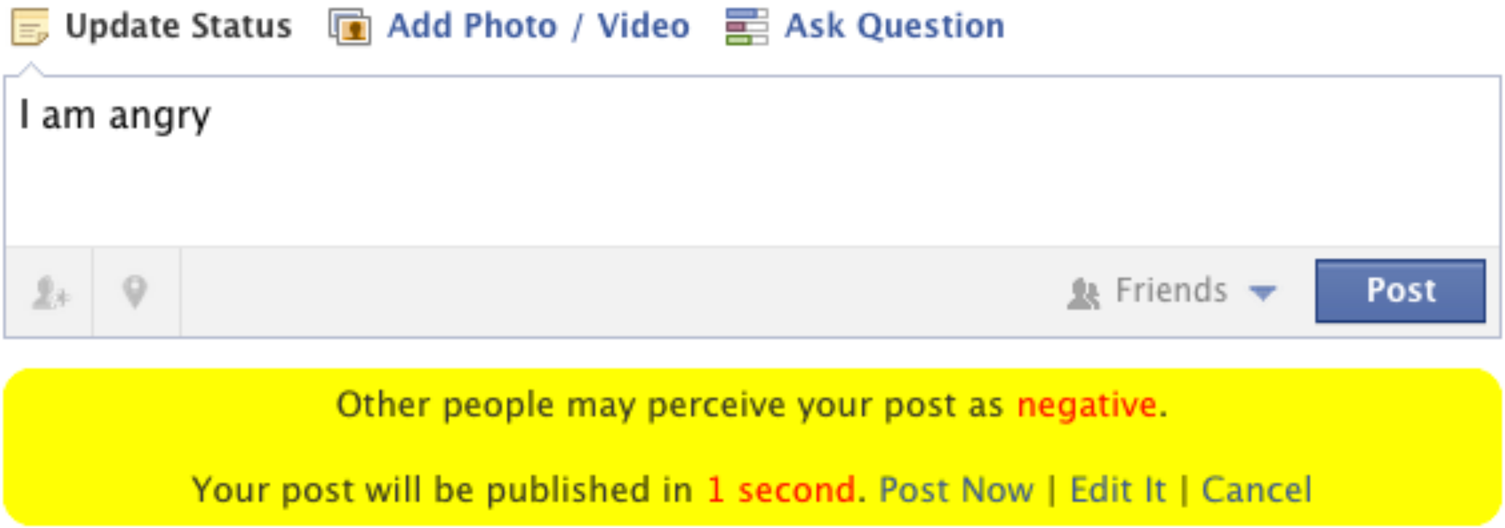
You will have 10 seconds to cancel after you post the update



heat in the moment

Your post will be published in 3 seconds. Post Now | Edit It | Cancel

Sentiment nudge (content feedback)



Update Status Add Photo / Video Ask Question

I am angry

Friends Post

Other people may perceive your post as negative.

Your post will be published in 1 second. Post Now | Edit It | Cancel

Social privacy technologies: challenges and limitations

- Focus on volitional actions and user-generated content
  - Limited by users' understanding and perception of the system
- Focus on the front-end
- Representativeness of user studies (mostly conducted in Europe and North America, mostly students)
- Focus on "privacy expectations"
  - Slippery slope if expectations erode
- Paradox of control (affects all types of privacy technologies)

- Incentives for deployment:
  - Aligned with industry's interests: make users comfortable with sharing information in their systems

“Institutional privacy”: Privacy concerns

- Interactions with organizations
- Data collection without user awareness
- Use of data for illegitimate purposes
- Sharing personal data with third parties
- Database breaches involving personal data
- Data correctness, integrity, deletion
- **Who** defines the privacy problem:
  - Legislation, organizations (through policies)

“Institutional privacy”: Goals

- Ensure compliance with data protection principles:
  - informed consent
  - purpose limitation
  - data minimization
  - subject access rights
- Data security:
  - prevent (or mitigate the consequences of) data breaches
  - protect user accounts
- Auditability and accountability

“Institutional privacy”: Examples

- appropriate defaults and privacy controls
  - opt-in vs opt-out
  - dashboards
- tools to make privacy policies easier to understand and negotiate
  - P3P, DNT
- tools to help organizations define and enforce access control policies
  - purpose-based access control
- auditing systems
- database security and privacy technologies

Institutional privacy technologies: challenges and limitations

- The organization is (semi-)**trusted** to be honest, competent, and act in the best interest of the user
  - Little or no (technical) protection if the organization wants to violate user privacy
  - Reliance on the legal system to punish lack of compliance
- Focus on limiting (mis)use of personal data, rather than collection
  - Does not preempt the creation of large databases
  - Auditing and legal compliance mechanisms may result in more data being recorded
- Who has the power to define and enforce the policies on data use?
  - Do whatever we wanted to do with the data while being compliant
- Focus on "personal data"
  - Does not address inferences from anonymized or aggregated data
- Limits on transparency posed by IP (proprietary software, algorithms, databases)

- Incentives for deployment: strong
  - Legal compliance is a very strong driver

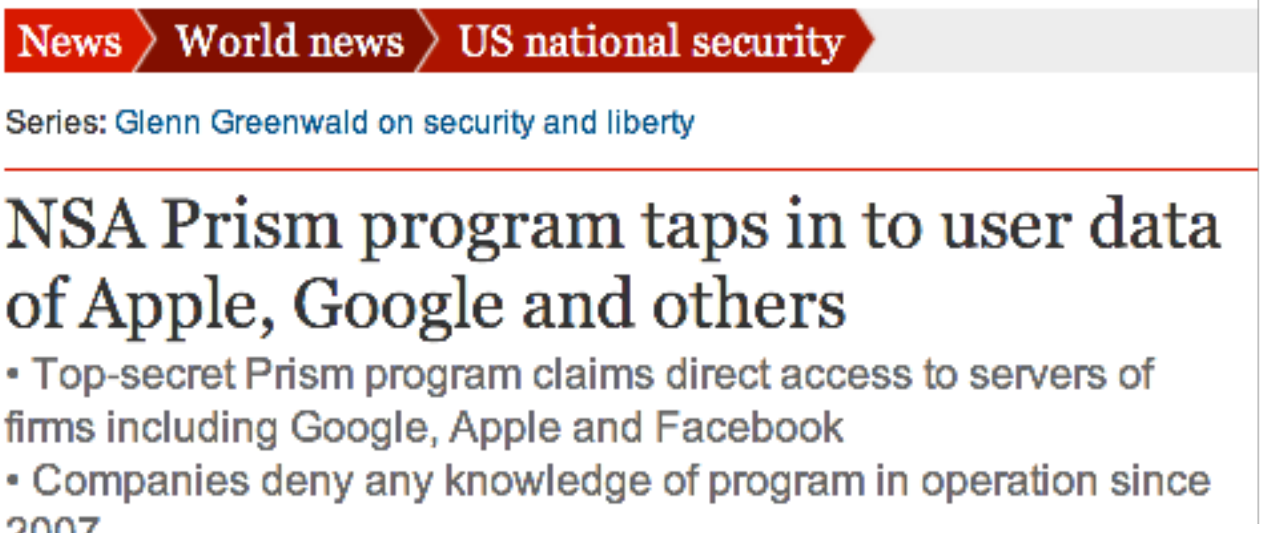
Anti-surveillance technologies (PETs): Privacy concerns

- Data disclosure by default through the use of the ICT infrastructure

- Surveillance by (possibly colluding) service providers and governments (abstract harms/consequences)

- Relationship to other democratic values:
  - Protection of dissent, free speech, freedom of association, freedom from government intrusion, protection of the democratic system itself

- **Who** defines the privacy problem:
  - Security experts (techno-centric)

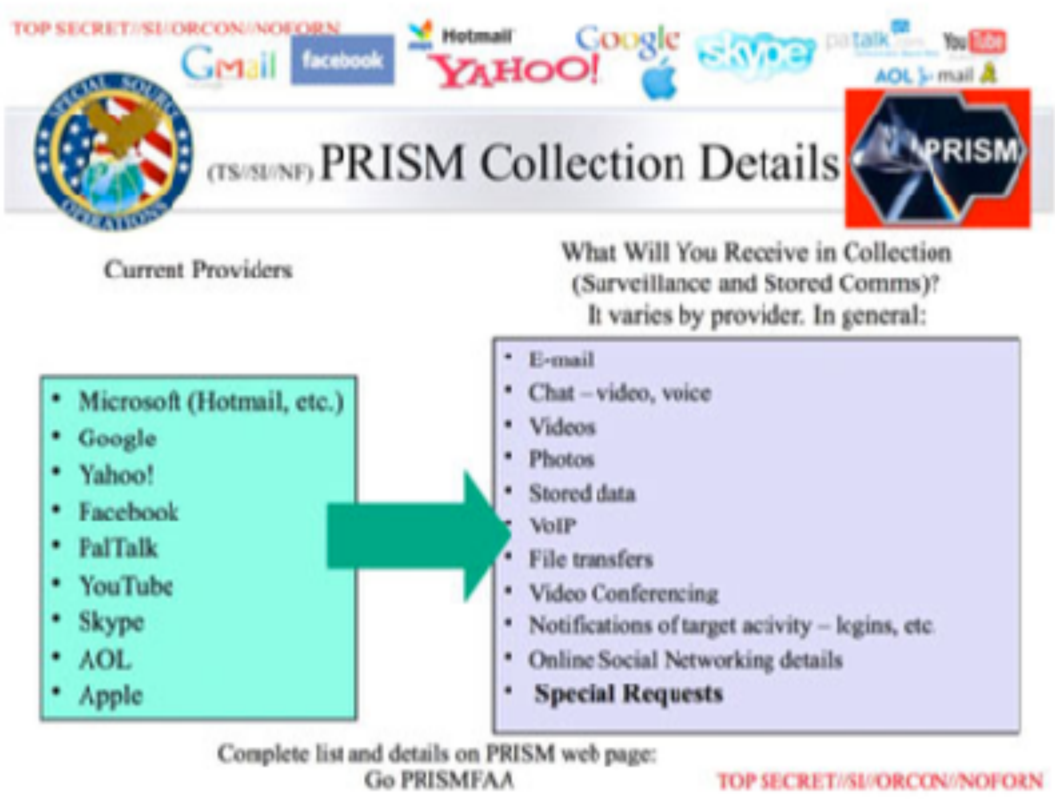


News › World news › US national security

Series: Glenn Greenwald on security and liberty

NSA Prism program taps in to user data of Apple, Google and others

- Top-secret Prism program claims direct access to servers of firms including Google, Apple and Facebook
- Companies deny any knowledge of program in operation since 2007



TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) PRISM Collection Details

Gmail
facebook
Hotmail
YAHOO!
Google
skype
paltalk
YouTube
AOL mail
PRISM

Current Providers

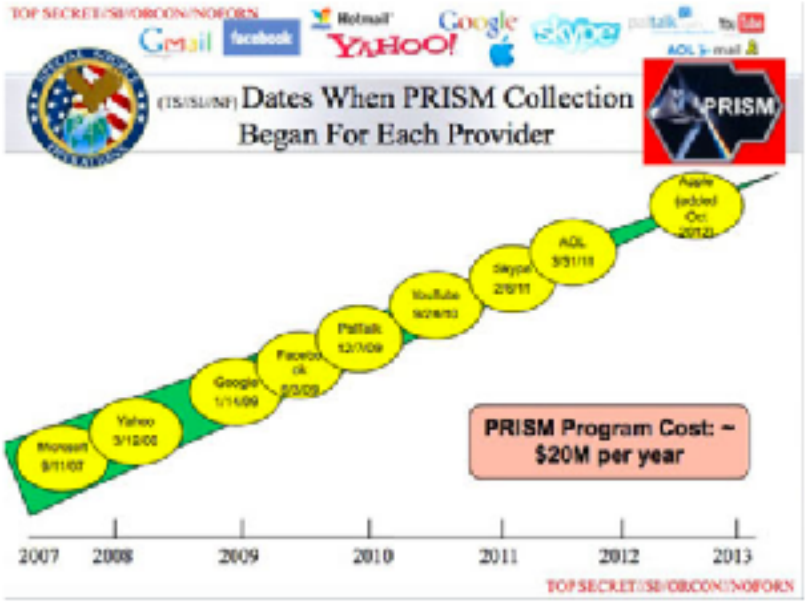
- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple

What Will You Receive in Collection (Surveillance and Stored Comms)?
It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target acivity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:
Go PRISMFAA

TOP SECRET//SI//ORCON//NOFORN



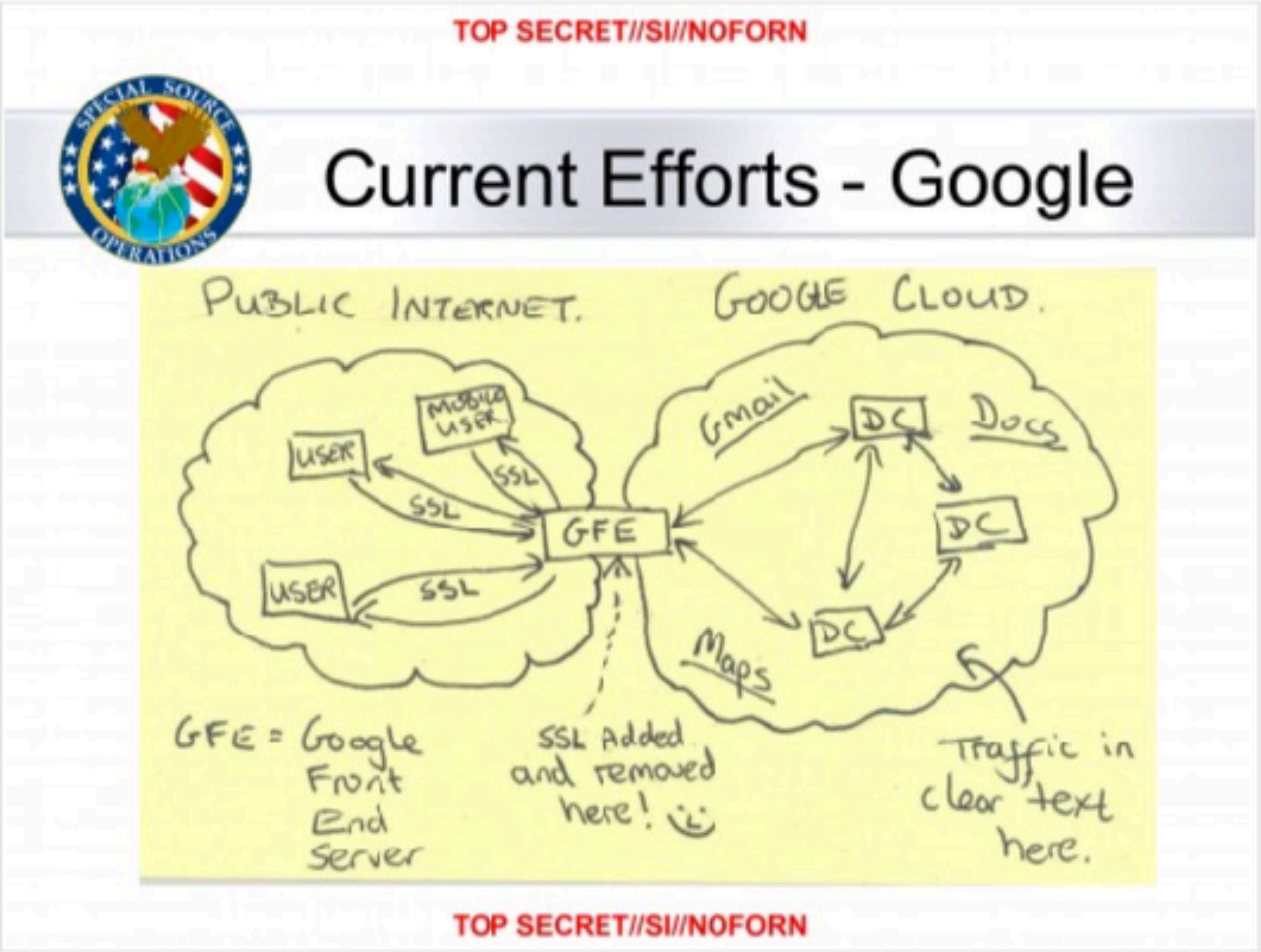
TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) Dates When PRISM Collection Began For Each Provider

PRISM Program Cost: ~ $20M per year

TOP SECRET//SI//ORCON//NOFORN

NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say



TOP SECRET//SI//NOFORN

Current Efforts - Google

PUBLIC INTERNET.
GOOGLE CLOUD.
USER
MOBILE USER
USER
SSL
SSL
SSL
GFE
Gmail
DC
Docs
DC
DC
Maps
GFE = Google Front End Server
SSL Added and removed here! :-)
Traffic in clear text here.

TOP SECRET//SI//NOFORN

In this slide from a National Security Agency presentation on "Google Cloud Exploitation," a sketch shows where the "Public Internet" meets the internal "Google Cloud" where user data resides. Two engineers with close ties to Google exploded in profanity when they saw the drawing.

Anti-surveillance technologies (PETs): Goals

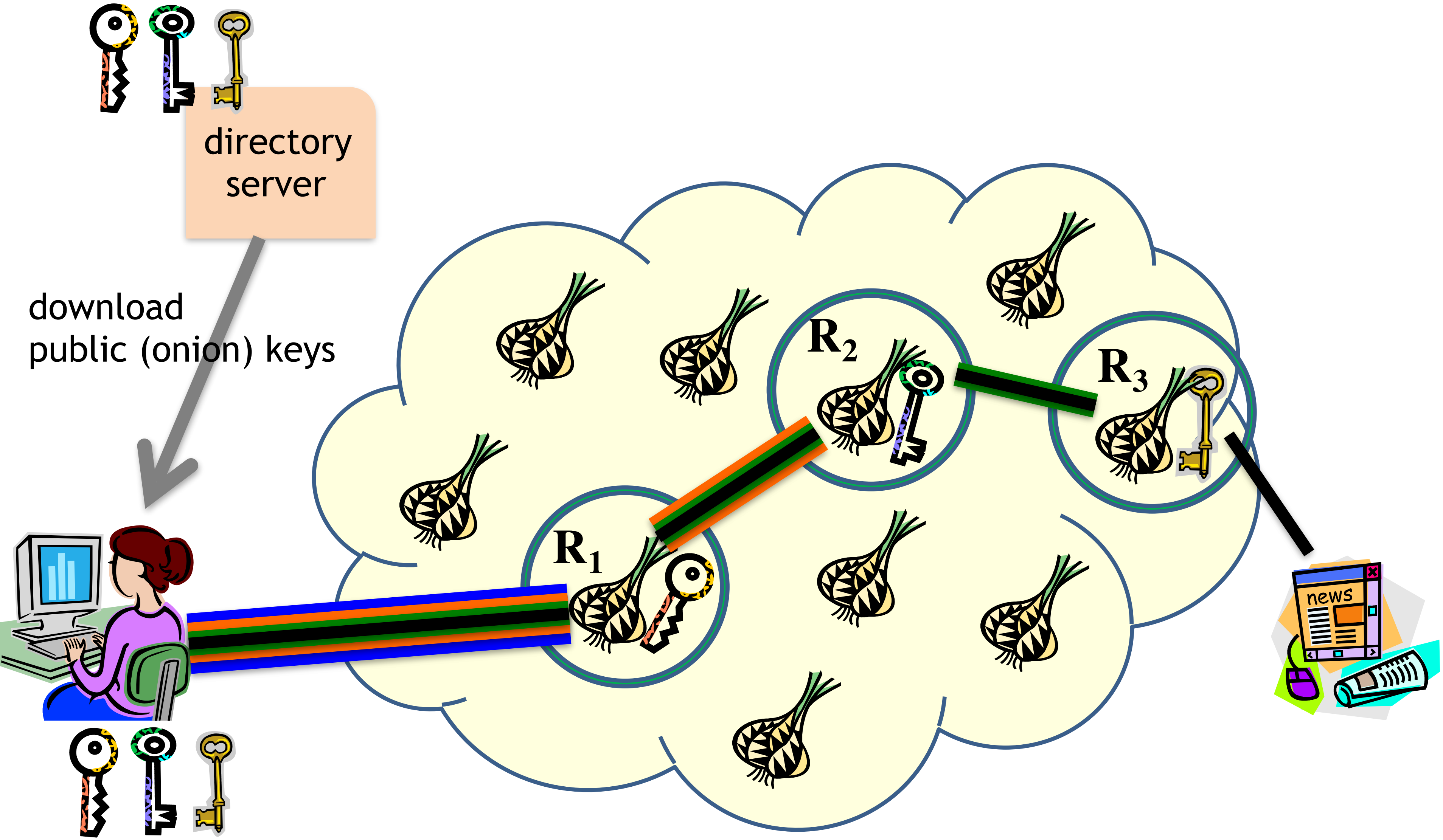
- Limit disclosure: prevent/minimize default **disclosure** of personal information to service providers and other third parties:
  - Only information *explicitly* disclosed is made available to *intended* recipients
  - This includes user-generated content *and* implicit data

- Minimize the **need to trust** others with appropriately handling data
  - Distribute trust by avoiding *single points of failure*
  - Transfer of trust to the technology (hard math problems, protocols, software, hardware) itself:
    - Need for **transparency**, availability of designs and implementations for public review

Anti-surveillance technologies (PETs): Examples

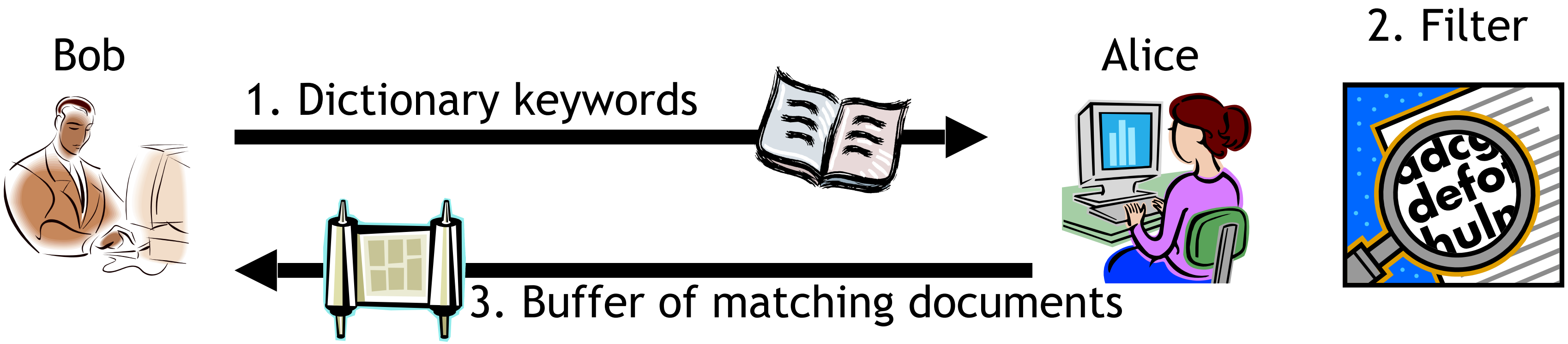
- Protecting content: end-to-end encryption
  - PGP, OTR
- Protecting identity: systems for anonymous communications
  - Tor
- Advanced crypto protocols:
  - anonymous authentication
  - private information retrieval
  - private search
  - privacy-preserving smart metering
- obfuscation approaches:
  - TMN, geo-indistinguishability, degrade data quality with noise
- Technologies that expose surveillance (transparency)
  - FPDetective

Tor



Private Search

- Alice stores documents
- Bob wants to retrieve documents matching some keywords
- Properties:
  - Bob gets documents containing the keywords
  - Alice does not learn Bob's keywords
  - Alice does not learn the results of the search



Bob

1. Dictionary keywords

Alice

2. Filter

3. Buffer of matching documents

Anti-surveillance technologies (PETs): challenges and limitations

- Focus on (preventing) data disclosure
  - No protection for information *after* disclosure

- Making secure design and implementations is hard
  - Many (hopefully explicit, sometimes implicit) assumptions need to hold to guarantee privacy properties.
  - Importance of public algorithms and open source: "it takes a village to keep systems secure"
  - Security of end-devices: big issue

- Narrow privacy definitions

- Making security usable is hard

- Incentives for deployment: weak at best

Conclusions

- Many valid ways of defining privacy

- Diverse landscape of privacy technologies, in terms of goals, limitations, and assumptions (trust, dependencies on technology, law, social norms or third parties)
  - hard to approach for outsiders (and even for insiders!)

- Importance of understanding embedded concepts of privacy and *who* gets to define those concepts and fill them with meaning!
  - keep some critical distance

- Privacy by Design
  - how to integrate the different technological approaches?

- Incentives!! Particularly, how to incentivize and support the deployment of anti-surveillance technologies?