Privacy and Security

Friends or Enemies?

JKU
JOHANNES KEPLER
UNIVERSITÄT LINZ

Michael Sonntag
Institute of Networks and Security

INSTITUTE
OF NETWORKS
AND SECURITY

Motivation

- "Privacy prevents security", "Data protection is offender protection", "You don't need privacy if you don't have something to hide"…
  - A common argument against privacy: You just cannot have it, as then security will be weakened
- But is this really true? How do they actually interact?
  - If security reports are only accepted with electronic signatures → Which employee will tell that there is a security issue/a successful attack has been kept secret/personal data was stolen/… ?
  - Encrypting communication (=privacy) also prevents modifications (=security) and identifies the server (=security)
  - Hiding which bank you use → Phishing gets more difficult
  - Using Tor for anonymity → Can save your life in some countries
- So what is the interdependence of privacy and security?
  - Can we have both or only one? How to do this?

Interdependence: Privacy needs Security

- Privacy: Keeping the associated person anonymous
  - Stored/Content data: Who this data is about
  - Communications: Who participates in the exchange
- Privacy needs security: Availability of data to everyone in cleartext means that privacy does not exist
  - This is more a problem than offline, as e.g. IP addresses always allow some tracing back, log files and metadata exists etc.
    - Paper: Cut letters from newspapers, use gloves, send by mail → anonymous communication
  - Big problem of IT security measures to improve privacy: Dependence on third parties
    - Anonymization systems require someone else to forward the message
    - Certification authorities know the identity of pseudonymous certificates
    - Bitcoin mixers must be trusted to not keep logs & dispense the “cash”
  - Solution (?): Chaining. Create chain; if at least one is trustworthy, anonymization works → Lots of security needed to ensure privacy

Interdependence: Security needs Privacy

- Security needs privacy:
  - Security researchers are sometimes attacked (e.g. DDoS on Krebs)
  - Whistleblowers provide security warnings – if they remain anonym.
  - Google indexes websites → Malware-Sites present a different view to Google (no malware) than to ordinary users (attacks)
    - Only if Google uses anonymous crawling this can be detected
    - Similar for all kinds of "undercover" investigations by the police
  - Uber presented fake information to government officials
    - Anonymous access would have allowed detection much earlier
  - No personal information stored → Much less desirable target
- But sometimes privacy is actually a problem:
  - Spam servers: Sender is anonymous → Spam filtering is weak
  - Online banking: Bank should know who you are
    - Note: Against third parties very important → Phishing mimics **your** bank!
  - Authorization: Requires identification
    - Potential solution: Bearer passes (Kerberos) → Ident. to third party only

Case study: IP addresses in logs?

- Legal case in Germany: May webserver logs store IP addresses (because of privacy this might be forbidden)?
  - Depends on whether IP addresses are personal data
    - ECJ: Yes, if the person storing them has the legal possibility of obtaining the identity of the person behind it
    - If possible only in case of an attack and through a court → Sufficient!
- Therefore IP addresses are personal data practically everywhere
- Result: Storage is only allowed if there exist overriding interests of the website operator, **e.g. because of security**
- Therefore this is a prime example:
  - More security → Less privacy: Store everything indefinitely
  - More privacy → Less security: Don’t store anything
- But is a third way possible?
  - Reduced storage for limited time only → Much better security and practically no privacy risk

Case study: IP addresses in logs?

- An expertise claims, IP addresses are not needed for security
  - Hashed values are sufficient if really needed
- But what if an attack is identified? If only the hash is available, the attacker can never be found → This is bad for privacy too, as stolen data can be used freely forever & the attacker hacks the next system!
  - **Too much privacy endangers privacy!**
- Also: IPv4 has only 4 billion addresses → Hashing is useless!
- What are the examples of privacy dangers?
  - Someone hacks the computer and steals the log files
    - Who has then difficulties tracing them back, as he needs ISP data!
  - Company steals data from ISP to identify its users
  - Company tries to identify users (e.g. login) and attributes all collected information to him/her → Already happens through cookies
- GDPR: Pseudonymity recommended → Automatic for IP addresses

Case study: Whistleblowing systems

- In whistleblowing systems privacy is paramount; security only second
  - Practically this is difficult and needs lots of work by the user
- Can we combine them into new and added functionality? Yes!
- “Whistleblowing confirmation” for company-internal systems
  - You send a report and obtain evidence of reporting it
  - If there are legal “problems” later you can always disclose (and prove!) that you did notify management of the problem
    - If nothing came from it → nothing more could be expected from you
  - This might enhance the willingness to disclose issues
- Lots of different security elements needed to achieve privacy
  - Which might be revoked by the company, so they must be verifiably (by the user) active and only work “forward”
    - Revocation only affects future (→ detectable), but not old disclosures
- If there is no privacy, the system becomes useless

Case study: Whistleblowing systems

- How to implement this?
  - The confirmation needs to contain the report
    - Or at least the full report must remain available in exactly the same form to the person reporting
  - Publicly registered signature from company, so it can't "vanish"
  - The confirmation may only be received after successful sending
  - The receipt of the confirmation must be ensured after sending
    - To be implemented through simultaneous disclosure protocols
  - Storing the document in a secure manner, so even in case of a search it is not found; but should survive accidents/fires/…
    - Like swiss bank accounts: They exist, but you (could not) get any information which accounts a specific person owns
    - Other options: Encrypted hidden containers, steganography, re-digitizable printout in secure storage …
  - Timestamp from third party to prove date and time of disclosure
    - And to prevent managers from retroactively creating disclosures, too!

Feature interaction

- How can we better understand and solve this relation?
  - “Feature interaction” is concerned with two (or more) features, which are perfectly fine alone, but lead to unintended consequences if they are present in the same system
    - Emergent behaviour because of interactions between the features
- While security and privacy are not “features”, they must be implemented through these (e.g. “file encryption” or “anonymization”)
- Feature interaction is a common privacy problem: “reidentification”
  - Datasets A&B are anonymous → together they identify the persons
- Centralisation might be a solution, as all interactions can be checked
  - But who implements this? Also a prime security & privacy target!
- Full autonomy would also be a solution: No interaction → no problem
  - But this is not what users want, as then e.g. a smart home only consists of a collection of smart devices, but is no “home”!

Feature interaction: Security and Privacy

- Deciding on the degree of centralization:
  - Compartmentalization: Centralization, but only for a limited area (not necessarily a physical boundary)
    - Reduces the amount of interaction and limits spreading personal data
    - Guideline: Use existing metaphors (“house”, “family”, “company”…)
  - Hierarchy: Interaction only with “neighbours”
    - Data and commands only to direct neighbours; more levels away only if data/commands have been “worked on” or are “aggregated”
      - If you can’t, pass it to someone who can; but no chain of sending on
  - Independence: Perform as much work as possible on your own, as less interaction with other systems eases analysis
    - Do not use the cloud if you can do it locally (even if this means more computing power is necessary); also helps if the cloud is unreachable, no longer provided etc. Prevents spreading personal data, as you don’t know whom the cloud will pass it along to

Case study: Script inclusion

- Feature interaction example: Websites including JavaScript libraries
  - Directly included → Possibly an outdated version
  - Indirectly included → Much more likely (partially double!) outdated
- Security implications:
  - Direct inclusion allows easier checks for updates
  - Allows verification what is included and whether it is unchanged
  - No double inclusion, no potentially conflicting versions
- Privacy implications:
  - No third party receives information on who visited which website
    - Example: Austrian newspaper “Die Presse” directly includes content from 13 other domains; even more if counting indirect inclusions!
- Result: Example where security **and** privacy benefit both through a single measure: Consolidate and put everything on your own server

Summary and outlook

- Privacy is a subordinate to security: No privacy without security, but security without privacy is technically possible
  - In many cases both can work together, and sometimes they must
- For both a clear definition is needed: Who is to benefit from them?
  - Security for the company only or also for users?
  - Privacy against third parties or also against the service provider?
- An integrated view is necessary, and privacy must be an important part of security, exactly as confidentiality, integrity etc.
  - “CIA” should be extended to “CIAP”, already in teaching
  - Privacy laws require some security → Security laws should also include privacy requirements
  - GDPR: Stronger focus on “privacy by design” is easiest to comply with by integrating it in a security analysis
    - And if such is currently lacking → privacy is an incentive to do it!

THANK YOU FOR YOUR ATTENTION!

Michael Sonntag
michael.sonntag@ins.jku.at
+43 (732) 2468 - 4137
S3 235 (Science park 3, 2nd floor)

**JOHANNES KEPLER
UNIVERSITÄT LINZ**
Altenberger Straße 69
4040 Linz, Österreich
www.jku.at