# Intro to privacy in computing (technical and legal privacy controls)

## Cos'è la privacy?

- The right to be let alone (essere lasciati soli)
- The right to individual autonomy (autonomia individuale)
- The right to a private life (vita privata)
- The right to control information about oneself (controllo delle informazioni di se stesso)
- The right to limit accessibility (limitare accessibilità)
- The right to minimize intrusiveness (minimizzare invadenza)
- The right to secrecy (segretezza)
- The right to enjoy solitude (godere la solitudine)
- The right to enjoy intimacy (godere l'intimita)
- The right to enjoy anonymity (godere anonimato)

La privacy è un diritto della persona, in realtà questo diritto oggi è abbastanza minacciato.

> Il concetto della privacy come diritto nasce nel 1890 da due avvocati **Warren & Brandeis** negli USA per una causa che riguardava un attore dove sono state pubblicate delle foto senza il suo permesso.

> Recent inventions and business methods call attention to the next step which must be taken for the protection of person, **and for securing to the individual what Judge Cooley calls the right "to be let alone"**.

> "Le fotografie istantanee e giornali hanno invaso il recinto della vita privata".

Il diritto evolve col il tempo con evolvere delle tecnologie, nel 2004, **Stefano Rodotà** il primo garante della privacy italiano, dice che in realtà **la privacy va un pò in là del diritto di essere lasciato solo, la privacy ha un legame profondo tra la libertà, dignità e privacy**.

> S. Rodotà dice che la possibilità di dare la forma che vogliamo alla nostra vita passa attraverso il controllo delle informazioni che ci riguardano, della nostra immagine, di ciò che vogliamo tenere per noi e di ciò che voglio che sia pubblico. La nostra stessa integrità è basata sulla possibilità di separare i piani della nostra vita, di assumere certi ruoli nella vita sociale più allargata ed altri nel privato, dove tendiamo a scoprirci di più e mettiamo più a rischio la nostra immagine. I progressi tecnologici hanno reso questa esigenza ancora pressante.

> Senza una forte tutela dell'informazione che le riguardano, le persone rischiano di più d'essere discriminate per le loro opinioni, credenze religiose, condizione di salute: la privacy si presenta così come un elemento fondamentale della società dell'eguaglianza. Senza una forte tutela dei dati riguardanti le convinzioni politiche o l'appartenenza a partiti, sindacati, associazioni, i cittadini rischiano d'essere esclusi dai processi democratici: così la privacy diventa una condizione essenziale per essere inclusi nella società della partecipazione.

> Senza una forte tute del "corpo elettronico", dell'insieme delle informazioni raccolte sul nostro conto, la stessa libertà personale è in pericolo diventa così evidente che: **la privacy è uno strumento**

> **necessario per difendere la società della libertà, e per opporsi alle spinte verso la costruzione di una società della soverglianza, della classificazione, della selezione sociale...**

Da questo momento la privacy evolve dalle fotografie personali **diventando un problema di tutelare le informazioni**, quindi, di sicurezza delle informazioni.

Garantire la confidenzialità delle informazioni significa garantire la privacy.

> La privacy diventa un diritto sociale e un valore sociale. (A.F. Westin)

Secondo Westin, se la privacy di una persona dovesse essere violata, quella persona **perde di autonomia**. Quindi, diventa ricattabile, quindi non è più autonoma nelle sue decisioni.

Una persona diventa ricattabile nel momento in cui qualcuno ottiene delle informazioni nei suoi confronti che non vuole che siano divulgate.

> Autonomy: "The most serious threat to individual's autonomy is the possibility that someone may penetrate the inner zone and learn his ultimate secrets, either by physical or psychological means. This deliberate penetration of the individual's protective shell, his psychological armor, would leave him naked to ridicule and shame and would put him under the control of those who knew his secrets."

> Self evaluation and decision making: Solitude and the opportunity for the reflection are essential to creativity. If every conversation among organizational leaders, if every draft memo, if every proposal for action were public, frank discussion would be severely inhibited and thoughtful decision making undermined.

Quindi, anche la capacità di decisionale è ricattabile.

Quindi, la privacy è **il diritto di una persona di potersi fare o di poter aver delle sfere private che condivide solo con le persone che vuole**. Che **equivale alla definizione di confidenzialità**, che **a certi tipi di informazioni possono accedere solo le persone autorizzate**, quando un sistema garantisce questo, rispetta il principio della confidenzialità.

La privacy può andare in conflitto con altri importanti valori della società:

- Prevenire e punire il crimine,
- La lotta contro il terrorismo,
- Facilità la diffusione delle informazioni false.

## Il valore economico delle informazioni

Un ulteriore elemento che è andato a peggiorare con avvento della tecnologia digitale è il valore acquisito delle informazioni per attività di marketing, **l'informazione non è più un bene personale ma sono diventate d'interesse di agenzie pubblicitarie**. Sapere quali sono gusti di una persona, quali sono le propensioni, quali sono le cose gli piacciono fare, può portare a **marketing mirato, quindi, indurre una persona a comprare**. Quindi, **le informazioni acquisiscono un valore commerciale**.

- Business intelligence market,
- E-commerce growth: influenzando le decisioni dei clienti si prevede un aumento di vendite di 6 trilioni di dollari,

- Impact of Data-driven decision-making: le decisioni guidate da informazioni, si aspettano che siano 5 volte più veloci rispetto la concorrenza.

L'informazione da una parte ha un valore sentimentale, dall'altra hanno un valore economico significativo.

Il concetto di privacy viene declinato in termini di informazioni.

## Digital Privacy

> The claim of individuals, group and institutions to determine for themselves, when, how and to what extent information about them is communicated to others. Alan Westin, Columbia University, 1967.

Il diritto di individui, gruppi e d'istituzioni di determinare come ed a quale punto l'informazione che riguarda può essere comunicata ad altri.

La privacy che ci interezza come informatici è garantire in qualche modo alle persone, alle istituzioni oppure alle organizzazioni su come e fino a che punto l'informazione che immettono nel sistema può essere comunicato agli altri.

In termini di privacy ci sono 4 tipi di informazioni:

- Personal Identifiable Information - **PII**: permette **l'identificazione immediata di una persona**. Es. indirizzo, numero della carta di credito, nome, numero di telefono, ecc.
- **NON-PII**: informazioni non sufficienti per identificare una persona ma **possono indicare tendenze e inclinazioni di una persona**. Es. preference personali (libri, film, musica, cibo), sesso, ecc.
- **Location Data**: (dati di locazione) non sono sufficienti per identificare una persona. es. coordinate gps.
- **Device/Network Data**, informazioni relative ai dispositivi. es. ip, mac, addresses, mail address, ecc. Non sono sufficienti per identificare una persona.

### Personal information - PII:

- nome cognome,
- data di nascita,
- codice fiscale,
- indirizzo, numero di telefono, passaporto, patente: informazioni sensibili,
- informazioni bancarie: informazioni altamente sensibili, usati per rubare l'identità degli individui.
- dati clinici, dati biometrici, dati estremamente sensibili.
- password e pin, cruciali per la sicurezza online,
- dati che riguardano il lavoro,
- dati accademici,
- documenti legali.

### Non-PII informazioni usati nella profilazione online

- siti visitati: sono informazioni sensibili che rivelano le preferenze e potenzialmente il comportamento online.
- search queries: informazioni che cerca online,
- comunicazione online: messaggi, conversazioni online,
- file scaricati e caricati,
- steaming e media consumption: musica e video che guarda online,

- attività su social media,
- transazioni online, e-commerce,
- location information,
- device information,
- application usage,
- VoIP calls.

Il problema di profiling è che tutte queste informazioni sono possedute da aziende come google, amazon e facebook.

# Privacy Threats

Quali sono le minacce alla privacy?

Minaccia alla privacy è una qualunque circostanza che può mettere a rischio la capacità di un individuo oppure di una entità a mantenere il controllo sulle informazioni personali e confidenziali, prevenendo accesso, uso, rivelazione oppure manipolazione non autorizzato.

Le minacce possono manifestare in molte forme come cyberattacks, data breaches, identity theft oppure altre attività che compromettono confidenzialità, integrità oppure la disponibilità delle informazioni sensibili.

Queste minacce possono sorge dalle vulnerabilità tecnologiche, dalle misure di sicurezza in adeguate oppure sfruttamento delle vulnerabilità del comportamento umano.

Una query a google è una minaccia alla privacy, se i sistemi di google vengono attaccati è un problema.

# Privacy attacks

I principali obiettivi di un attacco alla privacy sono:

- furto d'identità,
- profilazione,
- data breach.

## Identity theft

Furto d'identità, obiettivo è quello di impersonificare un'altra persona.

Attacchi mirate per rubare l'identita: phishing attacks, data breaches, social engineering, dumpster diving, skimming dives, malware and hacking.

Impatto sulla vittima: danni economici, danni all'imagine, problemi legali, ecc.

Scenari ad alto rischio: Wi-Fi pubblici, utilizzo di reti o connessioni non sicure, password deboli, ecc.

## User Profiling

Un tipo attacco mirato per costruire il profilo comportamentale di una persona, può essere di tipo psicologico oppure quello demografico. Utilizzato principalmente per il marketing mirato.

I dati utilizzati normalmente provengono da attività online (siti visitati, ricerche, ecc.), da attività sul social media, acquisti precedenti, ecc.

Tecniche usate per fare profiling sono: cookies, cross site scripting, tracking pixels, machine learning, data mining, social network analysis, data breaches.

Ovviamente, la maggior parte delle persone non sono consapevoli dei rischi, quindi, vogliono reti veloci, usabilità massima, usare password meno possibile oppure quanto meno usano password deboli, che la tecnologia gli seguisse in ogni momento e tutti sono pronti a usare l'ultima tecnologia a scapito della sicurezza.

Non realizzano che quando si utilizza un servizio quasi sempre il cliente e il prodotto è la persona stessa.

> "There's no such thing as a free lunch!"

- Se il servizio è gratuito, comunemente, l'azienda guadagna vendendo i dati dei propri clienti.
- Anche pagando un servizio, l'azienda potrebbe vendere i dati raccolti per un maggiore profitto.

**The Issue - Il problema**

Finché mostra qualche pubblicità di nostro interesse, potremmo anche tollerare. Ma cosa succede se ho il cancro e se questa informazione viene viene raccolto da ISP ed arriva all'assicurazione di conseguenza l'assicurazione ci alza il prezzo della polizza?

La società della sorveglianza: tutte le nostre attività finanziarie sono monitorati dalle istituzioni finanziarie per "rilevare frodi" oppure dai programmi fedeltà come quelle dei supermercati, i dati raccolti saranno utilizzati per marketing mirati.

Il cellulare può tracciare tutti nostri movimenti e comunicazioni "per servizi di emergenza".

La sorveglianza è qualcosa che può conferire accesso ai diritti e benefici ma allo stesso tempo può essere usato in maniera pericoloso, oppressivo e discriminatorio.

Il pericolo è che il potere della sorveglianza può diventare onnipresente.

I dati possono attraversare silenziosamente i confini delle nazioni, così che l'impatto della sorveglianza diventi difficili da identificare, regolare e dibattere.

è importante che questo potere, basato sulla supervisione delle attività e delle informazioni personali, sia esercitato in modo equo, responsabile, nel rispetto dei diritti umani, delle libertà civili e della legge.

## Data Breach

Data breach furto di informazioni sensibili. è l'accesso, la rivelazione, l'acquisizione non autorizzato alle informazioni personali che può provocare ad un potenziale danno, abuso o sfruttamento delle informazioni.

Cause comuni: cyberattacks (malware, hacking, ransomware), insider threats (negligenza dei dipendenti, dipendenti malintenzionati), furto fisico, ecc.

Data breach più famoso è quello fatto da Snowden.

Altri casi famosi:

- Equifax credit breach, dati di 150 milioni di persone
- Cambridge analytica, dati di oltre 87 milioni di persone

# Protections

Gli approcci sono di due tipi legale, tecnologico:

- **Legale**: si tratta di governare in qualche modo il traffico ed acquisizione di dati, il problema è che la legge arriva molto dopo la tecnologia, quindi, **non è in grado di prevenire**. Osserva delle situazioni e cerca di governarla. Quindi, ci sono leggi che sono state promulgate per contenere gli abusi di questi dati.
- L'approccio degli USA prevede un autoregolazione **Self-regulation**, cioè, ce un problema di abusi delle informazioni, tu, l'azienda fai del tuo meglio per evitarlo.
- **Tecnologico**: cercano di **ridurre il più possibile la circolazione** di questi informazioni. **Privacy Enhancing technologies PETs** sono tecnologie mirate per garantire la privacy, es. GPG.
- Educazione degli utenti, la cosa più difficile.

## Legal protection

Ormai è assodato che la privacy sia un diritto umano, quindi, ogni paese ha nella sua legislazione il riconoscimento della privacy come tale.

In europa la privacy è scritto nella carta dei diritti dei cittadini europei.

Ci sono due tipi di leggi,

- leggi universali (Comprehensive Laws): leggi generali che governa la raccolta, uso e disseminazione dei dati personali sia dal settore pubblico che privato.
- leggi settoriali (Sectorial Laws) come negli USA, ogni stato ha una sua idea diversa.

## General Data Protection Regulation - GDPR

La legge che regola la privacy a livello europea, è una direttiva, quindi, obbligatoriamente adottato da tutti gli stati membri UE. Entrato in vigore dal 25 Maggio 2018. Prevede delle sanzioni e previsto anche il carcere.

- The right to access (diritto accesso): sapere quali informazioni ha su di te.
- The right to be forgotten (essere dimenticati): posso chiedere di cancellare tutti dati in loro possesso.
- The right to data portability
- The right to be informed
- The right to have information corrected
- The right to restrict processing
- The right to object
- The right to be notified

GDPR ritiene personale anche i cookie e gli indirizzi IP, quindi anche un xss è ritenuto una violazione.

Le organizzazioni devono avere consenso scritto per raccogliere e a cosa vengono usate le informazioni.

Il consenso non deve essere ambiguo, può essere revocato e si può chiedere di cancellare i dati raccolti.

La raccolta delle informazioni non è vietata, ma deve essere preventivamente autorizzata da utente.

L'organizzazione deve garantire che le informazioni sono protette con adeguate misure di sicurezza contro data breach. In caso di data breach organizzazione è obbligata per legge informare dell'avvenuto data breach

(auto denuncia) e in questo caso il garante verifica se sono state adottate adeguate misure di sicurezza.

## Technical Privacy Controls - PETs

Cosa possiamo fare noi?

Dal punto di vista tecnologico l'unica cosa che può essere fatta è introdurre dei meccanismi che garantiscono:

- **Anonimato**: problema: la criminalità, un attacco ransomware utilizza questo fatti di anonimato per fare l'attacco.
- **Pseudonimità** Pseudonymity: l'identità di una persona è nota **solo in alcuni contesti** (soluzione alla anonimato), potrebbe essere anche falsa, quindi, ci sono essere abusi di varia natura.
- **Unobservability** e **Unlinkability**: in cui si riesce ad utilizzare una risorsa ma di fatto **non viene loggata** nessuna informazione.

### Depersonalizzazione

Il problema legato ai basi di dati che raccolgono i nostri dati, come fascicolo elettronico, un data breach su questo può compromettere una serie di aspetti.

Come facciamo che un data breach non comprometta la privacy di persone?

**Anonimizzare i database**, quindi, per esempio invece di memorizzare i referti con nome e cognome utilizzo un codice casuale e da qualche altra parte tengo una mappatura tra nome cognome e il codice. è la depersonalizzazione dei dati. Bisogna capire fino che punto possiamo depersonalizzare.

Un altro esempio è l'uso di DB cifrati (crypto database), è anche suggerito da GDPR. Ma il problema è quello di lavorare con DB cifrati.

# Lab

## Privacy Enhancing Technologies - PETs

PETs è un termine che racchiude un insieme di tecnologie che mirano ad aumentare il livello di privacy che include comunicazione sulla rete anonimamente, affrontare la sorveglianza, tracciamento dello preferenze, ecc.

Tools:

- End-to-End encryption
- GPG
- Tor Project

### Trusted Execution Environment - TEE

TEE è un ambiente "computazione sicuro" dove il codice viene processato isolatamente e protetta in termini di confidenzialità e integrità (nel senso che nessuno può accedere ai dati e nessuno può cambiare il codice e il suo comportamento).

Applicazione viene eseguito in una modalità del processore privilegiata.

Es. Netflix/Spotify utilizza TEE per processare i dati. Banche, Google pay, Apple pay.

 TEE

## Cookies

 Cookie Profiling

## Proxy for anonymous communication

Durante la comunicazione tra client e server, ISP può osservare tutti i dati che passa in chiaro. Un proxy ci permette di anonimizzare la connessione, lo fa praticamente facendo da tramite. Fondamentalmente, è un server tra il punto A e B. Quando ci colleghiamo ad un sito, la richiesta che fa il nostro browser viene inviato al proxy, proxy inoltra la richiesta e ci invia la risposta.

Il problema è il **single point of failure**, cioè, se il proxy è stato compromesso, può osservare tutto.

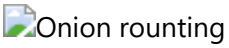
 Proxy

## TOR Project

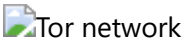
Tor basa sul concetto di proxy utilizzando diversi proxy (default 3) cifrando il messaggio a strati come una cipolla da cui arriva il nome `Onion Routing`.

Tor è un protocollo sviluppato da D.A.R.P.A. e US Navy. Invia il messaggio cifrato a strati attraverso una serie di server, il quale aumenta la difficoltà del tracciamento, aumentando l'anonimato.

Come funziona? Inizialmente, ce una fase di uno scambio delle chiavi tra vari nodi (chiamati anche `relay`). quando il messaggio arriva node1, il node1 decifra con la sua chiave ed inoltra al prossimo nodo e cosi via.

 Onion rounting

L'ultima parte rimane in chiara, quindi, bisogna usare HTTPS.

 Tor network

**Attacchi**

Tor attacchi di correlazione, in cui il primo nodo e l'ultimo nodo vengono attacchi per ricostruire il flusso di dati.