

Segurança da Informação

Márcio Moretto Ribeiro

10 de Agosto de 2017

Apresentação

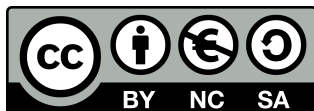
Essas são notas de aula da disciplina Segurança da Informação ministrados no segundo semestre de 2017 para as turmas do período diurno e noturno do curso de Sistemas de Informação da Escola de Artes Ciências e Humanidades (EACH) da USP. A primeira versão desta apostila foi escrita para o curso de verão ministrado entre os dias 2 e 6 de fevereiro de 2015 também no campus leste da Universidade de São Paulo. O curso de verão foi oferecido como parte das atividades do projeto de Privacidade e Vigilância do Grupo de Políticas Públicas em Acesso à Informação (GPoPAI) e foi inspirado pelo curso online oferecido gratuitamente pela plataforma Coursera e ministrado pelo professor D. Boneh.

Aos alunos que pretendem se aprofundar no tema sugerimos as seguintes referências bibliográficas:

- J. Katz e Y. Lindell - *Introduction to Modern Cryptography*
- W. Stallings - *Criptografia e Segurança da Informação*
- C. Paar e J. Pelzl - *Understanding Cryptography*

Agradecemos aos alunos que participaram do curso de verão em 2015 e dos cursos de graduação em 2016 e 2017, suas contribuições serviram de importante feedback para escrita dessas notas.

Alguns direitos sobre o conteúdo desta apostila são protegidos pelo autor sob licença Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). Ou seja, você é livre para distribuir cópias e adaptar este trabalho desde que mantenha a mesma licença, dê o devido crédito ao autor e não faça uso comercial.



Capítulo 1

Introdução

1.1 Privacidade

As primeiras tentativas de conceitualizar a privacidade datam do final do século XIX. Em um famoso artigo de 1890, os colegas da faculdade de direito de Bosto, Louis Brandeis e Samuel Warren denunciam que o advento da fotografia instantânea e o jornalismo haviam “invadido o recinto sagrado da vida privada” ferindo o que eles apontam como o direito individual de “ser deixado em paz”. Assim, eles argumentam que o escopo do direito comum (*Common Law*), que originalmente se concentrava apenas na proteção contra agressões e já havia sido ampliado para incluir a proteção à propriedade deveria ser novamente alargado para finalmente reconhecer a “natureza espiritual do homem” [WB90].

Essa visão da privacidade como proteção à intimidade ou, nas palavras dos autores, como o “direito de ser deixado em paz” é a chave de interpretação que prevalece no debate público. Duas contribuições que não fogem a esse espírito, porém, merecem destaque nessas notas introdutórias. Em um influente artigo de 1977, Richard Posner propõe que as disputas sobre direito à privacidade sejam interpretadas em sua dimensão econômica. Para Posner, as informações privadas possuem valor. De um lado há o interesse de quem busca construir uma imagem pública sobre si, uma reputação e, de outro, há o interesse de se conhecer o outro para melhor saber como se relacionar com ele ou aprender sobre sua trajetória. Caso fosse permitido qualquer tipo de intrusão à privacidade, o efeito seria um esforço maior em não revelar, ou inclusive não produzir, dados potencialmente valiosos. Assim, a questão

da privacidade, se resumiria a uma questão de eficiência econômica. Caberia ao legislador regular o direito ao controle da reputação procurando um equilíbrio que maximizasse os fluxos de informação [Pos77].

O risco tanto para o indivíduo cuja paz é perturbada pelas fotos não autorizadas, quanto para aquele que perde o controle sobre sua reputação é de que algo que pertencia a sua esfera privada se tornasse pública. Para Nissenbaum, essa dicotomização entre público e privado não dá conta dos problemas associados a quebra de privacidade. Por exemplo, um paciente espera que as informações sobre sua condição de saúde sejam eventualmente compartilhadas com outros médicos ou médicas com o intuito de melhor diagnosticá-lo, assim como um cliente espera que seu gerente de banco use suas informações bancárias para sugerir-lhe melhores investimentos. Porém, há uma flagrante quebra de privacidade se as informações médicas forem compartilhadas com o banco, com quem eventualmente o paciente negociará um plano de saúde. Esse cenário exemplifica o que a autora chama de rompimento da “integridade contextual do fluxo de informações” [Nis09].

1.2 Vigilância

Associado ao tema da privacidade, mas ligado a outra matriz teórica, estão os debates sobre vigilância. Diferente dos estudos sobre privacidade cujos principais autores são juristas preocupados com o direito individual, os estudos sobre vigilância focam em relações de poder. Foucault descreve a vigilância como uma técnica que teria alterado profundamente as formas de exercer o poder durante os séculos XVII e XIX. O poder do senhor feudal durante a idade média era exercido por meio do suplício, a pena corporal em que o açoitado pedia misericórdia eventualmente concedida. Após a revolução francesa o suplício foi sendo substituído pela prisão e aos poucos seria desenvolvida a técnica da disciplina e da vigilância. Para o autor, a imagem que melhor descreve a técnica é uma estrutura arquitetônica proposta por Jeremy Bentham no final do século XVII. Bentham arquitetou um modelo de prisão em que os vigias ficariam no centro aonde poderiam observar todas as celas, porém, aqueles que ocupam as celas não poderiam observar o vigia. A sensação constante de estar sendo vigiado introjetaria a disciplina, outra técnica deste período, nos condenados. O propósito da vigilância e da disciplina é o de produzir corpos dóceis e obedientes [Fou96].

Em 1992, em um curto texto, Giles Deleuze propos uma atualização dos

conceitos de Foucault que antecipariam o que hoje compreendemos como vigilância. Na sociedade disciplinar, descrita por Foucault, durante a vida o indivíduo passa de uma instituição disciplinar a outra: da escola, ao exército, do exército à fábrica e da fábrica ao hospital. Cada instituição disciplina o indivíduo e o modela da maneira mais eficiente à instituição. Na sociedade do controle, conforme descrita por Deleuze, o poder é exercido de maneira mais intermitente e mais sutil. O indivíduo prototípico da sociedade do controle seria o endividado cujo controle atravessa as instituições [Del92].

1.3 Marco Regulatório

Antes ainda dos primeiros computadores, as chamadas máquinas Hollerith revolucionaram a capacidade de processamento de dados. Durante a década de trinta elas dinamizaram o processamento dos dados do censo nos EUA e na década de 40 foram usadas pelos nazistas para classificar aqueles, principalmente judeus, mas também comunistas e homossexuais, que deveriam ser transportados para os guetos, dos guetos para os campos de concentração e finalmente para as câmaras de gás [Bla01]. Finda a guerra, a evolução dos modernos Estados de bem estar social Europeu e seu necessário processamento massivo de dados casou muito bem com o desenvolvimento computacional e assustou os cidadãos com sua centralidade de processamento. Assim, começaram a surgir as primeiras leis de proteção de dados pessoais.

Mayer-Schonberger argumenta que, uma vez que as leis de proteção de dados pessoais na Europa partem todas das mesmas bases e diferem apenas em detalhes, é mais frutífero estudá-las em conjunto do que seguindo uma análise comparativa. Ele propõe uma abordagem geracional como se existisse uma tendência evolutiva das normas. A primeira geração, no começo dos anos 70, focou na regulamentação técnica dessas bases centralizadas de dados. O surgimento de mini-computadores, que favorecia o processamento descentralizado, levou a uma adaptação na legislação. A segunda geração, no final dos anos 70, focou na liberdade negativa, o direito civil de "ser deixado em paz" nas palavras de Brandeis e Warren. A autonomia do indivíduo é, porém, contraposta a sua inclusão nos programas sociais do Estado. Então, a terceira geração legislativa, em meados dos anos 80, foge um pouco das liberdades negativas e foca em uma abordagem participativa de autodeterminação informacional. A pergunta deixa de ser se alguém quer participar ou não de processos sociais, mas como. Ainda assim, porém, os indivíduos

estavam em uma posição frágil nas relações de negociação o que os levava, via de regra, a abdicar dos seus direitos. A quarta geração, de meados dos anos 90, procurou de um lado equalizar as posições de negociação ainda apostando na autonomia do indivíduo, mas também incluiu diversos mecanismos mais paternalistas excluindo certas liberdades participativas e as sujeitando à proteção jurídica obrigatória. Nessa fase surgem órgãos de defesa, não apenas de auxílio aos cidadãos, mas com papel decisório para deliberar contra violações [MS97].

No Brasil, o Marco Civil da Internet aprovado em 2013 não aborda diretamente as questões de proteção de dados pessoais. Carecemos de um marco legal que imponha, pelo menos, que o uso de dados pessoais dependa necessariamente do consentimento explícito e informado e cuja autorização seja dada para um fim específico.

1.4 Vigilância Digital em Massa

Em 2013 Edward Snowden revelou ao mundo o alcance dos programas de vigilância em massa das agências de espionagem dos EUA. O jornalista Glen Greenwald e a cineasta Laura Poitras divulgaram o caso em uma série de matérias e um documentário [Gre14, Poi14]. O vazamento demonstra que a agência de segurança nacional dos EUA (NSA) tem acesso a toda a comunicação por telefone e pelos principais meios de comunicação online do mundo. O moderno modelo de negócios das empresas de internet baseado na propaganda direcionada depende da construção de perfis digitais que por sua vez dependem da produção e aquisição de uma grande escala de dados pessoais. Essa competição por dados pessoais cria o que chamamos de pontos únicos de falha. A violação dessas bases permitiu à NSA produzir um banco de dados pesquisável da agência possui toda a comunicação pública e privada que passa pelos servidores da Google, do Facebook, da Microsoft e da Apple.

A vigilância digital em massa eleva o problema da privacidade para um outro patamar. Não se trata apenas de proteger a intimidade, ou a inviolabilidade do lar, ou do controle na construção da reputação. Nesse contexto, o problema da privacidade é também coletivo. A privacidade deve ser também encherada como um direito civil, uma limitação ao poder do estado de antecipar as ações de grupos políticos. Para tanto, é preciso de ação política de conscientização, de regulamentação para restringir o poder das empresas que controlam o armazenamento dos dados pessoais e também desenvolvimento

técnico.

1.5 Segurança da Informação

A internet é um meio intrinsecamente promíscuo []. Por uma decisão de projeto, não temos controle por onde nossas informações passam quando nos comunicamos pela rede. Conforme produzimos mais informações pessoais e permitimos que elas circulem, maior o risco de quebra da integridade dos fluxos contextuais. Em particular há atores poderosos com capacidade conhecida de observar a comunicação em escala global o que traz um risco coletivo tanto à soberania nacional dos países periféricos, como o Brasil, quanto à democracia. A regulamentação, absolutamente necessária para controlar minimamente esses processos e garantir pelo menos o consentimento no uso de nossas informações pessoais, certamente não é suficiente. A compreensão, o desenvolvimento e a difusão de ferramentas de segurança da informação podem colaborar nesse sentido. Concluiremos o capítulo com uma história motivadora.

Após as denúncias de Snowden houve uma espécie de consenso nos meios ativistas sobre a importância de focar forças em desenvolver ferramentas que garantissem a criptografia ponta a ponta. O paradigma mais comum de comunicação na rede é criptografar a comunicação entre cada cliente e o servidor. Como já dissemos, conforme poucos servidores consentiram a maior parte da comunicação online, a informação armazenada nesses servidores passa a ser um bem muito requisitado. A ideia para superar isso seria criptografar a comunicação entre clientes. Assim, a informação armazenada nos servidores não seria compreensível seja pelos engenheiros das empresas que controla a comunicação, seja para um ator externo como um hacker ou a NSA. O principal protocolo de criptografia ponta a ponta na época era o PGP, que havia sido criado no começo da década de 90, antes do advento da web. As tentativas mal sucedidas de ressucitar o protocolo logo foram substituídas por um esforço em atualizá-lo. Duas aplicações que garantiam criptografia ponta a ponta em celulares se popularizaram nesse período: Telegram e o Textsecure. A primeira foi desenvolvida por uma companhia russa e oferece serviço de criptografia ponta a ponta em comunicação síncrona usando um protocolo desenvolvido por seus engenheiros. A segunda foi desenvolvida por uma pequena empresa no Vale do Silício e se inspirou no protocolo OTR, que por sua vez se inspirou no PGP, adaptando-o para o contexto assíncrono mais

adequado para a comunicação móvel. Os esforços de ativistas em promover esse tipo de ferramenta culminou com a adoção do protocolo do Textsecure, rebatizado como Signal, no Whatsapp, a ferramenta de comunicação móvel mais usada no mundo todo. A popularização da criptografia ponta a ponta em grande parte da comunicação interpessoal muda muito o cenário de proteção de direitos civis e de liberdade de organização. É certo que os metadados das comunicações - quem fala com quem, quando e de onde - não estão protegidos, é certo que a maior parte da comunicação interpessoal não está livre de intrusão seja de hackers, seja de agências governamentais, é certo que há serviços - como agenda online - em que simplesmente não há alternativa segura e, portanto, é necessária muita ação política e desenvolvimento técnico nessa área.

Capítulo 2

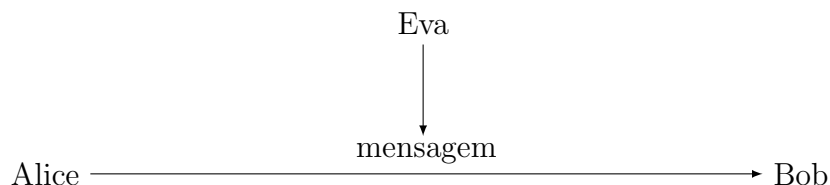
Cifras Clássicas

Como argumntamos no primeiro capítulo, a internet é um meio de comunicação promíscuo. As partes que se comunicam pela rede não tem controle sobre por quais caminhos sua comunicação irá trafegar. Essa característica, porém, não se restringe a esse meio. Durante o século XVIII, por exemplo, toda correspondência que passava pelo serviço de correios de Viena na Áustria era encaminhada para um escritório – *black chamber* – que derretia o selo, copiava seu conteúdo, recolocava o selo e reincaminhava para o destinatário. Todo esse processo durava cerca de três horas para não atrasar a entrega. Como a Áustria, todas as potências européias desse período operavam suas *back-chambers*. As invenções do telegrafo e do rádio só facilitaram a capacidade de criar grampos, no primeiro caso, ou simplesmente captar a comunicação no segundo.

Partiremos, portanto, do seguinte modelo de comunicação. Duas partes, o remetente e o destinatário, buscam se comunicar. Tradicionalmente denominaremos o remetente de Alice e o destinatário de Bob. Nossa suposição principal é que o canal de comunicação entre as partes é inseguro. Ou seja, assumiremos que terceiros, que denominaremos de Eva, são capazes de observar as mensagens que trafegam pelo canal de comunicação. Essa suposição é conhecida em alguns meios como “hipótese da comunicação hacker”. Para efeitos deste curso, sempre assumiremos essa hipótese.

A *criptografia* (do grego “escrita secreta”) é a pratica e o estudo de técnicas de comunicação segura na presença de terceiros chamados de *adversários*. A *criptoanálise*, por sua vez, é o estudo e a análise dos sistemas de informação com a intenção de desvelar seus segredos. Nosso primeiro desafio no curso é apresentar sistemas de comunicação que garantam a *con-*

fidencialidade. Ou seja, toda mensagem enviada de Alice para Bob deve ser compreensível apenas para Alice e Bob e deve ser incompreensível para Eva:



Se a importância da comunicação confidencial entre civis tem se tornado cada vez mais urgente, no meio militar é difícil remontar suas origens. Suetônio (69 - 141) por volta de dois mil anos atrás descreveu como o imperador Júlio César (100 a.c. - 44 a.c.) escrevia mensagens confidenciais:

“Se ele tinha qualquer coisa confidencial a dizer, ele escrevia cifrado, isto é, mudando a ordem das letras do alfabeto, para que nenhuma palavra pudesse ser compreendida. Se alguém deseja decifrar a mensagem e entender seu significado, deve substituir a quarta letra do alfabeto, a saber 'D', por 'A', e assim por diante com as outras.”

O esquema que chamaremos de cifra de César é ilustrado pelo seguinte exemplo:

Mensagem: `transparenciapublicaopacidadeprivada`

Cifra: `XUDQVSDUHQFLDSXEOLFDRSDFLGDGHSULYDGD`

Como descrito Suetônio, a regra para encriptar uma mensagem consiste em substituir cada letra da mensagem por aquela que está três posições a sua frente na ordem alfabética. Para descriptografar a cifra, substituir cada letra por aquela que está três posições atrás. O problema com este tipo de sistema é que basta conhecer a regra de criptografia para decifrá-lo. Em outras palavras, o segredo da cifra é sua própria regra.

Embora técnicas de criptografia e criptoanálise existam desde o império romano, foi com o advento do telégrafo e sua capacidade de comunicação eficiente, que o campo se estruturou. No fim do século XIX Auguste Kerckhoff estabeleceu seis princípios que as cifras militares deveriam satisfazer:

1. O sistema deve ser indecifrável, se não matematicamente, pelo menos na prática.

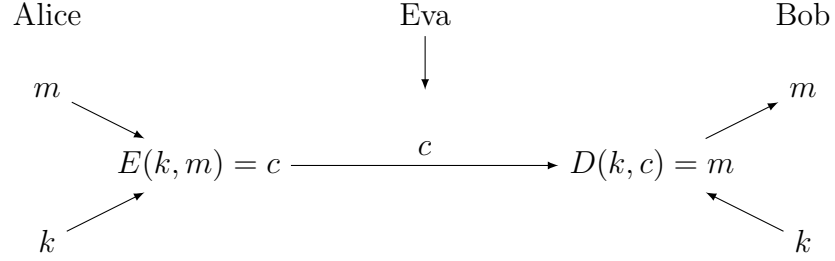
2. O aparato não deve requerer sigilo e não deve ser um problema se ele cair nas mãos dos inimigos.
3. Deve ser possível memorizar uma chave sem ter que anotá-la e deve ser possível modificá-la se necessário.
4. Deve ser possível aplicar a sistemas telegráficos.
5. O aparato deve ser portátil e não deve necessitar de muitas pessoas para manipulá-lo e operá-lo.
6. Por fim, dadas as circunstâncias em que ele será usado, o sistema deve ser fácil de usar e não deve ser estressante usá-lo e não deve exigir que o usuário conheça e siga uma longa lista de regras.

O segundo princípio ficou conhecido como *princípio de Kerckhoff*. Ele estabelece que a regra usada para criptografar uma mensagem, mesmo que essa regra esteja codificada em um mecanismo, não deve ser um segredo e não deve ser um problema caso ela caia nas mãos do adversário. Nas palavras de Claude Shannon: “o inimigo conhece o sistema”. Whitfield Diffie coloca o debate nos seguintes termos:

“Um segredo que não pode ser rapidamente modificado deve ser interpretado como uma vulnerabilidade”

Ou seja, em uma comunicação confidencial as partes devem compartilhar algo que deve ser “possível de modificar caso necessário”. Esse segredo compartilhado é o que chamaremos de *chave* da comunicação e assumiremos que ela é a única parte sigilosa do sistema. Trazendo o debate para uma discussão mais moderna, o sigilo do código-fonte de um sistema não deve em hipótese alguma ser aquilo que garanta sua segurança.

O modelo de *criptografia simétrica*, portanto, pode ser descrito da seguinte maneira: o remetente usa um algoritmo público (E) que, dada uma chave (k), transforma uma mensagem (m) em um texto incompreensível chamado de *cifra* (c), a cifra é enviada para o destinatário por um meio assumidamente inseguro (hipótese da comunicação hacker) e o destinatário utiliza a mesma chave em um algoritmo (D) que recupera a mensagem a partir da cifra.



2.1 Cifra de Deslocamento

O que chamamos na seção anterior como “cifra de César” não deve ser propriamente considerado uma cifra, pois não possui uma chave. Porém, é possível e simples adaptar esse esquema para incorporar uma chave. Para tanto faremos a seguinte alteração no esquema. Ao invés de deslocar as letras sempre três casas para frente vamos assumir que foi sorteado previamente um número k entre 0 e 23. Esse número será a chave da comunicação e, portanto, assumiremos que as partes a compartilham. O mecanismo para criptografar uma mensagem será o de deslocar cada letra k posições para a direita e para descriptografá-la basta deslocar cada letra as mesmas k posições para a esquerda.

Para formalizar este mecanismo vamos assumir que cada letra do alfabeto seja representada por um número: a letra **a** será representada pelo 0, a letra **b** pelo 1 e assim por diante. O universo de todas as chaves possíveis é o conjunto $K = \{0 \dots 23\}$ (chamaremos este conjunto de \mathbb{Z}_{24} ou de maneira mais genérica $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$) e o universo de todas as mensagens possíveis é representado pelo conjunto $M = \mathbb{Z}_{24}^*$, ou seja, todas as sequências de números entre 0 e 23. Além disso, o conjunto das possíveis cifras é $C = M$. Precisamos descrever três algoritmos:

- Gen que gera a chave $k \in K$,
- E que recebe uma chave $k \in K$ e uma mensagem $m \in M$ e produz uma cifra $c \in C$ (i.e.: $E : K \times M \rightarrow C$) e
- D que recebe uma chave $k \in K$ e uma cifra $c \in C$ e produz uma mensagem $m \in M$ (i.e.; $D : K \times C \rightarrow M$).

Um sistema de criptografia simétrica Π é formado por essa tripla de algoritmos $\Pi = \langle Gen, E, D \rangle$. Além disso, precisamos garantir que quem possui

a chave seja capaz de descriptografar a cifra. Ou seja, precisamos garantir que:

$$D(k, E(k, m)) = m$$

O mecanismo que gera uma chave na cifra de substituição é bastante simples, ele simplesmente sorteia com uma distribuição de probabilidade uniforme um número entre 0 e 23. Escreveremos da seguinte forma:

$$Gen := k \leftarrow \mathbb{Z}_{24}$$

Utilizaremos a partir daqui a convenção de usar uma seta da direita para esquerda indicando que será escolhido um elemento do conjunto com probabilidade uniforme.

O algoritmo para criptografar uma mensagem traz um pequeno problema. Escreveremos $m = m_0m_1m_2 \dots m_n$ uma mensagem m com $n + 1$ letras cuja primeira letra é m_0 , a segunda é m_1 e assim por diante. Nossa primeira tentativa de formalizar E seria somar k a cada uma das letras m_i . O problema é que esta soma pode resultar em um valor que não corresponde a nenhuma letra i.e. $m_i + k > 23$. Para evitar este problema utilizaremos não a aritmética convencional, mas a *aritmética modular*.

Dizemos que um número a divide b (escrevemos $a|b$) se existe um número inteiro n tal que $a.n = b$. Dois números são equivalentes módulo n (escrevemos $a \equiv b(modn)$) se $n|(b-a)$. Em outras palavras, dois números são equivalentes módulo n se o resto da divisão de cada um por n for o mesmo resultado. O conjunto de todos os números equivalentes módulo n forma uma classe de equivalência que representaremos como $[a mod n] = \{b \in \mathbb{Z} : a \equiv b(modn)\}$. Por exemplo $[5 + 7 mod 10] = [2 mod 10]$ pois $5 + 7 = 12$ e o resto de 12 por 10 é 2.

Estamos finalmente em condições de formalizar o sistema da cifra de deslocamento $\Pi = \langle Gen, E, D \rangle$:

- $Gen := k \leftarrow \mathbb{Z}_{24}$
- $E(k, m) = [m_0 + k mod 24] \dots [m_n + k mod 24]$
- $D(k, c) = [c_0 - k mod 24] \dots [c_n - k mod 24]$

2.2 Cifra de Substituição

Bibliografia

- [Bla01] E. Black. *IBM e o Holocausto: a aliança estratégica entre a Alemanha nazista e a mais poderosa empresa americana*. Editora Campus, 2001.
- [Del92] Gilles Deleuze. Postscript on the societies of control. *October*, 59:3–7, 1992.
- [Fou96] M. Foucault. *Vigiar e punir: história da violência nas prisões*. Vozes, 1996.
- [Gre14] G. Greenwald. *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. Henry Holt and Company, 2014.
- [MS97] Viktor Mayer-Schönberger. Technology and privacy. chapter Generational Development of Data Protection in Europe, pages 219–241. MIT Press, Cambridge, MA, USA, 1997.
- [Nis09] H. Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books. Stanford University Press, 2009.
- [Poi14] Laura Poitras. Citizenfour. Documentary, 2014.
- [Pos77] Richard A. Posner. The right of privacy. *Sibley Lecture Series*, 22, 1977.
- [WB90] Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 4(5):193–220, December 1890.