Governance and regulation

Ross Anderson
The technology cycle

- Steam, steel, electricity, radios, cars, oil all followed a similar pattern:
  - First, ground-breaking investment in new tech
  - Second, speculative frenzy in an evolving market
  - Third, consolidation, correction, and regulation to ensure safety and correct market power
  - Fourth, a mature industry settles down

- The Internet is now entering the third phase

- We see the same cycle at smaller, faster scales for component technologies
Governance is complex!

• Global and local public goods
• Security and safety certification
• Consumer protection
• Regulatory capture
• Competition policy and antitrust
• Defending democracy
• Protecting the vulnerable
• Privacy versus surveillance
Global public goods

• Public goods are non-rival and non-excludable
• Examples: national defence, clean air
• Increasingly these are global not local! Clean air was smog; now, limiting CO$_2$ emissions
• Tech adds many more, complex and inter-related:
  • a dependable Internet (and local networks too)
  • tolerable levels of cybercrime and abuse
  • security standards and related safety standards
  • trust in commerce, and in governance itself...
The Orange Book

- 20\textsuperscript{th} century approach: the Trusted Computing Systems Evaluation Criteria
- MLS systems for sale to NATO governments
- Evaluation done by (US) civil servants
- Careful but took forever – government systems were always a generation out of date
- Small quantities meant they were unreasonably expensive
The Common Criteria

• Evolved from UK/F/DE/NL ITSEC system, and adopted widely from late 1990s
• Write a suitable Protection Profile, evaluate to it
• A big deal for smartcards, HSMs and other kit for banking, ID management, electronic signature
• All but highest levels of assurance delegated to commercial licensed evaluation facilities (CLEFs)
• Each country’s CLEFs are regulated by its national agency (GCHQ in Britain)
• Failure mode: vendors shop for the cheapest deal
Compliance regimes

• ISO 27000 process for documenting security management processes – basically run by the Big Four audit companies

• Healthcare systems in the USA have the HIPAA compliance regime

• Quoted US companies: Sarbanes-Oxley, etc

• Financial regulation: FCA in the UK for banks, firms offering credit; PSR for payments. Everything from crypto and resilience standards to anti-money-laundering and know-your-customer duties
Cyber Essentials

• UK government frustration with easy CC evaluations from formerly communist countries
• Also frustrated with auditor-driven processes like ISO 27000 – almost all hacked firms of any size are ISO 27000 certified
• Cyber essentials launched post-Brexit to provide a minimum baseline for government suppliers
• University issue: how to we certify that all devices are patched up to date, and still allow user devices? Are service expiry dates consistent with Apple?
Core problems of governance

• Everybody grabbed a share of security standard setting until it didn’t work any more
• There are more general failure mechanisms!
• People’s own interests aren’t the same as their employer, and firms’ objectives aren’t society’s
• We use laws to fix this, but laws are made by legislators who are human
• Powerful organisations lobby to change the rules in their favour...
Regulatory capture

• Regulators often end up run by ‘their’ industries
• The expertise comes from there!
  • FCA, MHRA
• Sometimes politicians design regulators to be weak
  • ICO
• Sometimes there’s arbitrage too
  • Ireland’s data protection commissioner
• Sometimes there’s deception
  • Security standards with backdoors for intel access
Competition policy

• Monopolies have come and gone in our industry: NCR, IBM, Microsoft, Google/Facebook...
• TikTok is now beating Google as the leading online destination (Dec 2021)!
• US/UK largely abandoned antitrust enforcement from the 1980s thanks to the consumer surplus test
• Monopoly is not just tech (see Matt Stoller’s blog)
• The EU has historically been stronger
• The USA is starting to change under Biden!
Defending democracy

• We can pass content moderation laws – the UK Online Safety Bill (going through parliament)
• Raises privacy issues – what should FB look at?
• And competition issues – FB can afford to hire another 15,000 moderators, but can a startup?
• What are the broader effects of legislating for mandatory content filters?
• How will such mechanisms end up being used in less democratic countries?
Protecting the vulnerable

• Banks try to blame customers for fraud – losers tend to be poor, women, minorities
• KYC and other ‘security theatre’ make transitions harder, e.g. escaping a partner, changing gender
• Assumption of mental capacity disadvantages the elderly, children
• But protecting kids / seniors properly is hard!
• Politicians talk a lot about child protection online; the Budapest Convention (2004) prohibited CSAM
Protecting children

• Beeban Kidron’s Age-Appropriate Design Code is now in force:
  • High level of privacy for under-18s by design and default
  • Don’t share location by default
  • Make location and other privacy settings obvious
  • Don’t nudge children to make harmful choices
  • Don’t auto-recommend harmful stuff
  • Turn off behavioural advertising...

• But child protection talk is often used to justify quite different policy goals
Privacy versus surveillance

• Claims about protecting kids or stopping terrorists used for years to justify surveillance powers

• 1990s: ‘Crypto war 1’ when US, UK governments tried to limit strong cryptography

• Outcome: lots of crypto today is weak, as with Mifare Classic, or has protocol issues, as with Bluetooth, or certification issues, as with TLS

• June 2020: EU announces demand for ‘client-side scanning’ of end-to-end encrypted apps

• Aug 2021: Apple announces a design
Existing content scanning systems

• Nazi material (F, De); terrorism (EU); child sex-abuse material (many); spam, animal cruelty, nudity

• Usually done on providers’ servers with mix of human moderators and tech:
  • Perceptual hashing (still images)
  • Machine learning (NLP, videos)

• Moderators help build target lists / training data for filter models

• Not very effective (FB gets 25% of hate speech in English but only 2% in Arabic)

• Expensive (FB has 15k moderators)
Threats to content scanning

• Abuse by authorized parties (e.g. Australian police raid journalists who publish war-crime photos)

• Scope creep, e.g. extending from child abuse to missing children by adding face recognition, then adding dissidents too

• Abuse by unauthorized second parties, e.g. corrupt police, tech company insiders

• Abuse by unauthorized third parties, e.g. foreign states and criminals

• Local adversaries, such as your partner, ex-partner or personal rival
Location of scanner

• If scanning is done in WhatsApp, move to Signal
• If in the device O/S, attacker gets everything (cloud forensics too)
• If in device middleware (Apple proposed the back mechanism for the iOS Camera Roll), opt-out may be possible in theory, hard in practice
• If kept at the server, can run much bigger models (e.g. video, NLP) and detect many attacks on the mechanism
Apple offer, Aug 2021

• Scan all photos when uploaded from iPhone’s Camera Roll to backup in iCloud

• NeuralHash, a perceptual hashing technique, checks each photo you take / import against a block list of 200,000 historical child sex abuse images

• Once 30 uploaded photos are on the block list from NCMEC, fancy crypto lets them be decrypted

• Apple staff / contractors review for possible false alarms, and report real abuse images to authority
Effects of moving scanning to client?

• Access to stored data, not just comms
• Reveal content other than legitimate targets (to both authorized and unauthorized abuser)
• Reveal content to local adversaries
• Reverse engineering of targeting material (reversible hashes, ML models’ training data)
• Attackers can experiment to improve attacks
• More software → more vulnerabilities
Effects of moving scanning to client (2)

- Evasion attacks on perceptual hashes get easier
- False-positive attacks may also be easier to devise (Apple’s NeuralHash had second-preimage attacks found within days)
- Adversarial machine-learning attacks on ML can be used for evasion, poisoning and backdooring (e.g. police to covert population-wide search for photos of Bin Laden / Dalai Lama / the Pope)
- Jurisdictional issues become harder
Academic response

• “Bugs in our Pockets: The Risks of Client-Side Scanning”
  • Hal Abelson, Ross Anderson, Steve Bellovin, Josh Benaloh, Matt Blaze, Jon Callas, Whit Diffie, Susan Landau, Peter Neumann, Ron Rivest, Jeff Schiller, Bruce Schneier, Vanessa Teague, Carmela Troncoso

• Client-side scanning extends bulk surveillance from device communications to storage

• Makes law-abiding citizens and whole societies more vulnerable

• But does not guarantee effective crimefighting
Wrapping up...

• Lots of stuff fails because of conflicting incentives both within and between organisations
• Governments try to fix things, but they have mixed incentives of their own
• There’s adversarial behaviour all the way up and down the stack!
• Expect a long hard journey on tech governance – as with other industries before us
• Meanwhile you need to study the power dynamics, so you know when you’re fighting the right battle...