One piece of the puzzle: Computational limits and opportunities of anonymity for whistleblower protection

Stefan Schiffner
University of Münster

Bettina Berendt
TU Berlin, Weizenbaum Institute, and KU Leuven

CIF Seminar, 1 March 2022

Where is whistleblowing happening?

Public Sector
- Admin (corruption)
- Secret Services
- Military
- Law enforcement

Private Sector
- Insider trading
- Creative book keeping
- Abuse of power

What makes these different?

Public Sector
- Preservation of power

Private Sector
- Monetary advantage
  - Company
  - Individual
Why is whistle-blowing important?

‘The Enron of Germany’: Wirecard scandal casts a shadow on corporate governance

PUBLISHED MON, JUN 29 2020 4:37 AM EDT | UPDATED MON, JUN 29 2020 5:22 AM EDT

Ryan Browne
@RYAN_BROWNE

• The Wirecard accounting scandal has raised fresh questions about corporate governance, with some experts calling it the “Enron of Germany.”

• German financial regulator BaFin has come under fire for its handling of the situation, with the government now calling for regulatory reform.

• There are also questions about why EY, Wirecard’s auditor, didn’t pick up on accounting irregularities that date back years.
Retaliation prevents whistleblowing; what encourages whistleblowing?

• The risk of retaliation is a major disincentive for potential whistleblowers (WBs).
• How to reduce this risk? Make retaliation
  • illegal (or at least protect WBs legally)
  • otherwise shunned or even unattractive
  • impossible
Whistleblowing as a communications problem (a very simplified view) (1)
Whistleblowing as a communications problem (2)

Audience

Wrongdoer

sees

Whistleblower

tells

responsible for,
bears consequences for, ...
Whistleblowing as a communications problem (3)

- **Audience**
- **Wrongdoer**
- **Whistleblower**

Whistleblower sees the wrongdoer responsible for, bears consequences for, ..., tells the audience who can retaliate against.
Whistleblowing as a communications problem (4)

Whistleblower

sees

Wrongdoer

tells

Audience

can retaliate against
Whistleblowing with anonymity: the promise

- **Whistleblower**
- **Anonymity set**
- **Audience**
- **Wrongdoer**

The diagram illustrates the concept of whistleblowing with anonymity. A whistleblower sends a message to the audience, who sees the whistleblower and the anonymity set. The wrongdoer, who can retaliate against the whistleblower, is also depicted in the diagram.
Threats to anonymity

- Mistaking *confidentiality* for anonymity
  - a trusted entity knows the identity, pressure on this entity can reveals the identity
- *Direct re-identification*: based on cues
  - legal name, pseudonyms, fingerprints/DNA, unwise choices case management data
- *Addresses* of various types
  - physical location, email address, telephone number, IP address, GPS coordinates
- *Security measures*: Need-to-know, tracking, logging
- Inferences from report *metadata*
  - e.g. when a report was made, the voice of the reporter on a telephone hotline, the linguistic style and revealed lingo of a written report
- *Epistemic non-anonymisability*: Who are the knowers?
  - Small anonymity set. The message content may imply identity.
Anonymity is hiding in the masses

Idea: using general anonymity services such as Tor to hide.

SecureDrop (originally DeadDrop by Aaron Swartz and Kevin Poulsen, 2013)
• Implemented as hidden service in Tor
• Target Userbase: Journalists and their sources
• NT, Intercept, Süddeutsche, apache.be

GlobaLeaks (2010)
• Implemented as Tor hidden Service
• Target Userbase: WB in Public service
• AWP: Ljost (Iceland), Filtrala (Spain), EcuadorTransparente, PeruLeaks
SecureDrop Architecture

Whistleblower: Sees something; tells something.

Journalist: Collects info from sources; reports to the public.

WB’s Internet-connected PC with Tor browser

Encrypts Message using: \( k_p \) (\( \bullet \) )

SecureDrop app. server

Journalist’s Internet Connected PC

Secure Viewing Station; generated key pair \((k_p, k_s)\)

air gap (floppy, USB stick, etc.)

Journalist: Collects info from sources; reports to the public.

Whistleblower: Sees something; tells something.
SecureDrop Administrative Domains

Whistleblower:
Sees something; tells something.

WB’s Internet-connected PC
with Tor browser

Encrypts Message using: $k_p(\cdot)$

SecureDrop
app. server

Journalist’s Internet
Connected PC

Journalist:
Collects info from sources; reports to the public.

Secure Viewing
Station; generated key pair ($k_p, k_s$)

air gap
(floppy, USB stick, etc.)
SecureDrop’s Protection Goals and Threat Model

Protection Goals
- Sender Anonymity
  - Towards everyone (NSA, Network Provider, Receiver)
- Confidentiality
  - WB can establish a confidential channel to the Journalist

Threat Model
- Attacker can observe:
  - What it can observe with Tor
  - The journalists Internet PC
- Attack cannot access
  - WB PC
  - Secure viewing station
  - SecureDrop server
Attacker can observe:
- What it can observe with Tor
- The journalist's Internet PC

- WB's Internet Connected PC with Tor browser
- Encrypts Message using: $k_p$ (key)
- SecureDrop app. server
- Attack cannot access WB PC Secure viewing station SecureDrop server
- Journalist's Internet Connected PC
- Secure Viewing Station
- Whistleblower
- Journalist
Case Study: Could SecureDrop have saved Reality Winner?

- US Airforce trained cryptologic linguist -> left Airforce for inner conflict, realizing her translations helped to kill people
- Hired by Pluribus as translator and assigned a job at US Army post “Fort Gordon”
- She stumbled upon documents that implicated Russian hacking attacks targeting the 2016 US presidential elections
- Assessing that the government would not act upon this intelligence, she decided to leak these documents to Intercept
Case Study: Could SecureDrop have saved Reality Winner?

- *Intercept* shared these documents with the NSA for verification.
- NSA started investigation: 6 suspects had access to said documents;
- only Winner contacted *Intercept* from her work computer
- Further documents appeared to be scanned from hard copy. Scans contained ID dots of the printer.
Case Study: Could SecureDrop have saved Reality Winner?

- Winner was charged with “removing classified information from a government facility” → 63-months sentence.
- She was released under probation in mid 2021.
Potential SecureDrop gain for Winner

• SecureDrop would have hidden that Winner contacted Intercept
  • However: initial anonymity set of 6 is small (other potential clues like timing very likely).

• Main evidence: Printer’s Machine Identification Code
Is the air gap in SecureDrop real?

- Depends on the actual machine for the SVS
  - Can it write on any other device?
  - Does it store any plaintext?
  - How is plaintext otherwise treated?
  - Who provides this machine?
- Case Winner: Journalist sent original evidence to NSA
Journalist’s burden even if technology is magically provided

• Winner’s anonymity set was reduced by the fine-grained evidence delivered by Intercept to the NSA.
• Journalist would need to know what NSA knows to assess what can be shared.
• There might be no utility left...
So far for journalists; what is different for our application

• Trust assumptions:
  • No “neutral” instance that can set up and run the system
Outlook: The anonymity set depends (also) on the wrongdoer – risk and opportunities

- Anonymity set: Whistleblower sees and tells about the wrongdoer.
- Wrongdoer & management can retaliate against the whistleblower.
- Audience can influence anonymity possibilities through technical and organisational measures.
Discussion:
Choices influencing anonymity possibilities?

• Fictitious ex. of management choices that trade off a security loss against a whistleblower-incentivisation gain: “k-anonymous access control” + ”k-anonymous logging”

• How could/would this be argued?

• In different domains?
  • “top-secret documents” (Winner) vs.
  • “accounting data” (Wirecard)