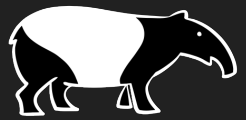
WHY?

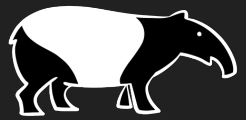
2024-11-26



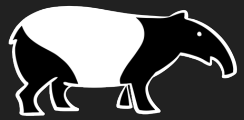
Theories in group communication strategies

say starting a question with **Why** triggers defensiveness

and should be avoided

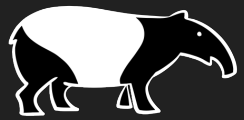


This is problematic



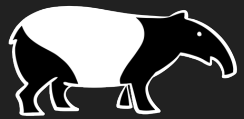
Most presentations of DNS TAPIR have focused on

What and How



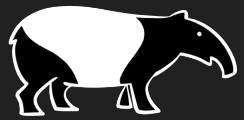
We're proud of **What** we have built and **How**

Alluding to **Why** with terms like *Privacy* and *Security*, we assume the ideas behind the project are apparent

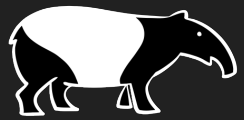


This assumption is false

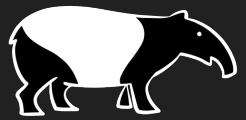
and a mistake I make frequently



Why matters



WHY ME?



Engineer and Internet geek

TIPnet + TeliaNet + others, Consultant, Nominum/Akamai

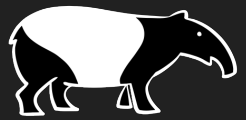
Hobbies: Political science, astronomy, philosophy, economic theory, psychology…

Former member of the Surveillance Industrial Complex

Ex techno-utopian

Ex techno-dystopian

Currently, founder of DNS TAPIR



Are we net positive?

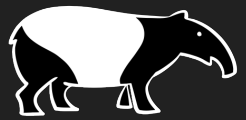
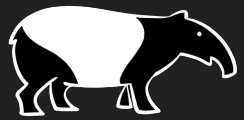


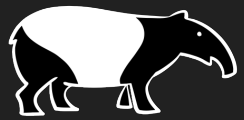
Image by winnifredxoxo

Is the degree of improved security sufficiently large to make the increased level of privacy invasion acceptable?

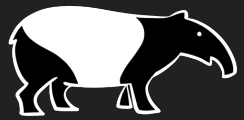
Problem I **DEGREE**	To what degree do we trade privacy vs security
Problem II **SECURITY**	How do you measure security?
Problem III **PRIVACY**	How do you value privacy?



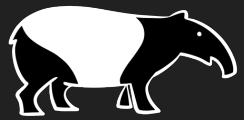
Why, though?



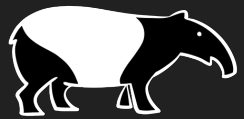
I want to work towards a society I want to live in



How many of you think it’s ok to have video surveillance outside the Swedish parliament?



How many of you think it’s ok to use facial recognition technology in the Stockholm subway system to catch fare dodgers?



“Like humans, facial recognition algorithms can accurately infer gender, age, ethnicity, or emotional state. Unfortunately, the list of personal attributes that can be inferred from the face extends well beyond those few obvious examples. A growing number of studies claim to demonstrate that people can make face-based judgments of honesty, personality, intelligence, sexual orientation, political orientation, and violent tendencies.”

Kosinski, M. Facial recognition technology can expose political orientation from naturalistic facial images. Sci Rep 11, 100 (2021). https:// doi.org/10.1038/s41598-020-79310-1

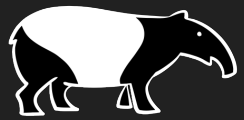


Image by Nick-K (Nikos Koutoulas), CC BY 2.0

Monitoring vs Surveillance

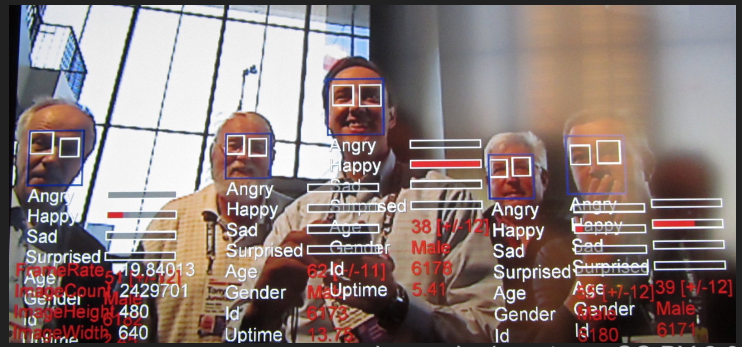


Image by jurvetson, CC BY 2.0

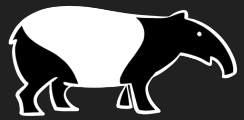
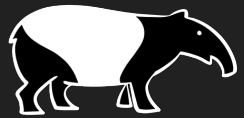


Image by Nick-K (Nikos Koutoulas), CC BY 2.0

These are not the same

Image by jurvetson, CC BY 2.0

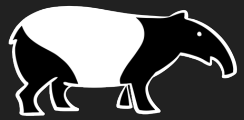


DNS TAPIR
Privacy first data collection and analysis
Anonymisation and differential privacy
Transparency in code and data
Independent, non-profit, “data commons”

DNS TAPIR vs Others

🐝 ☔ 🦉

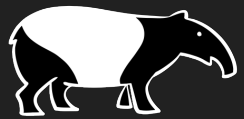
Typical Others
Big data centralized collection and analysis
Pseudonymisation and data protection
Proprietary
Varying goals and motivations



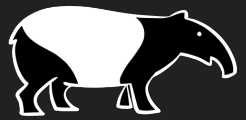
DNS TAPIR
Privacy first data collection and analysis
Anonymisation and differential privacy
Transparency in code and data
Independent, non-profit, “data commons”

These are also not the same

Typical Others
Big data centralized collection and analysis
Pseudonymisation and data protection
Proprietary
Varying goals and motivations



Precautionary principle



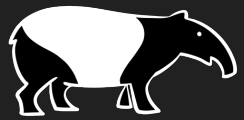
… applied to data collection:

- ❏ If you don’t need it, don’t collect it
- ❏ If you only need it briefly, don’t store it
- ❏ Analyse data, conscious of sensitivity and relevance
- ❏ Continuously evaluate reasons for retaining data

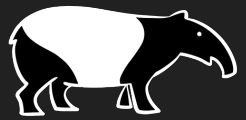
… with the following drawbacks:

- ❏ Incomplete or lacking information
- ❏ Incomplete or lacking context
- ❏ Inflexible analysis

The trade-off is a conscious choice



HyperLogLog



The problem

HLLs can be brute forced

Small groups may generate the exact same HLL

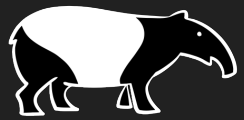
Solutions

Core intensive

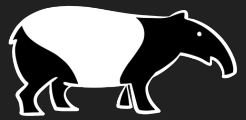
- ❏ Initialize each round of HLLs with 100 fake clients across all sketches
- ❏ In processing, remove the universally common group of clients

Edge intensive

- ❏ For rows with fewer than N clients, send numeric count
- ❏ Retain HLL for an hour
- ❏ If still below N clients, give numerical hourly count



DNS blocking



The problem

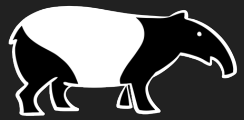
DNS blocking is done with intel with very little provenance

You implicitly trust vendors to be correct, unbiased and immune to manipulation

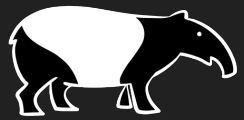
Solution

- ❏ We only create observations with provenance.
- ❏ To the policy processor (POP) we only provide notable observations.
- ❏ Don’t trust us, make up your own mind!

Hats off to **Quad9**, who address this problem doing post-hoc analysis to identify false positives (or mischief).



But really, Why?



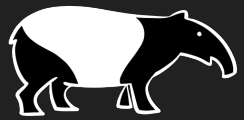
Current collection strategies are ethically questionable

Reversing the effects of surveillance is difficult

Amassing toxic datasets with dangerous alternate uses

Current data collectors say it can’t be done differently

That’s why we do it differently



Thanks!

Questions?