Metadata flow analysis

and privacy

STRINT Workshop, Ted Hardie presenting

If you have the data, you have the metadata

The more confidential the underlying data the more useful metadata protection is; protecting the envelope is not all that useful if the text of the greeting is still visible...

Possible mitigations

Using these may raise the costs of surveillance:

- **Aggregation**
- **Contraflow**
- **Multipath**

Aggregation

Pooling traffic makes it harder to identify who initiated any specific flow.

-HTTP Proxies, caching resolvers, TURN servers do this now.

-Content pooling also makes analysis more difficult

-Usefulness depends on where signal is gathered, but also creates a target.

Contraflow

Tunneling traffic to alternate exit points makes it more difficult to correlate with other traffic.

- Every VPN increases the cost of surveillance

- The larger and more varied the number of exit points, the more useful.

-Impacts peering and generally increases number of route miles consumed.

Multipath

Varying the link or path traffic takes also makes correlation harder.

-Split tunnel VPN approaches

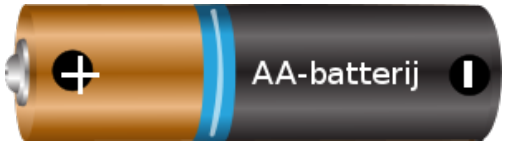
-Cellular/WiFi uplink selection

-Time-based selection of traffic origination ("Work Ted/Home Ted")

-Increases uplink costs, number of route miles, interface selection complexity

Design Considerations

- Don’t make your protocol automatically resist these mitigations.
  - Consider protected channels and app tokens for state
  - Don’t assume topology equals destiny (or geolocation)
  - Consider the impact of aggregation



- Designing mitigations
  - Combining these may be a better bang for the buck.
  - Consider how to avoid using these becoming a trigger
  - Nothing is perfect; raising the cost of pervasive surveillance is more important than the last 9