Designing Privacy into Internet Protocols

IAB Privacy Program

Why are we here?

- Security as an IETF design consideration (RFC 1543, 2223, 3552, 3365, …)
  - Realistically cannot design and standardize a new protocol without confidentiality, authentication, integrity, etc. protections or strong story for why not.
- RFC 6973 extends these considerations to privacy and formalizes them.
- Today's goal:
  - Walk out with some idea of how to incorporation privacy considerations into protocol design and motivation to learn more.

Scope

- Narrow: focused on individuals.
- Broad: any information relating to an individual who can be identified, directly or indirectly, may be relevant.
- Limits to what can be addressed in protocol design (vs. deployment and operation).
- No explicit prohibitions or requirements.
- Distinction between (negative) defending against exploits and (positive) building privacy tools.
- Discussion without reference to any particular legal framework.

Disclaimers

- Nature of communicating is that you reveal some data. Communicating without revealing anything at all is very difficult.
- Adding privacy or security protections in one area can reduce privacy in others.
- Remember that very few perfect solutions exist.

Agenda

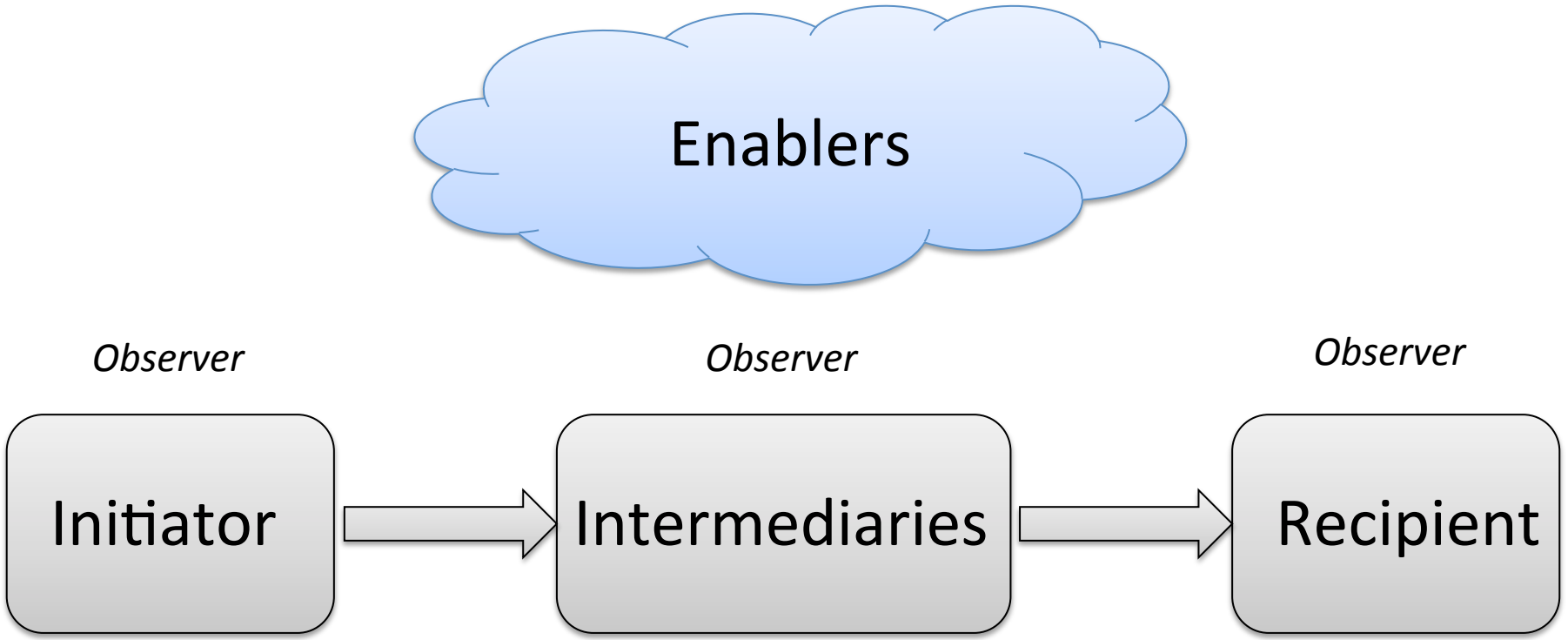
Comm. Model

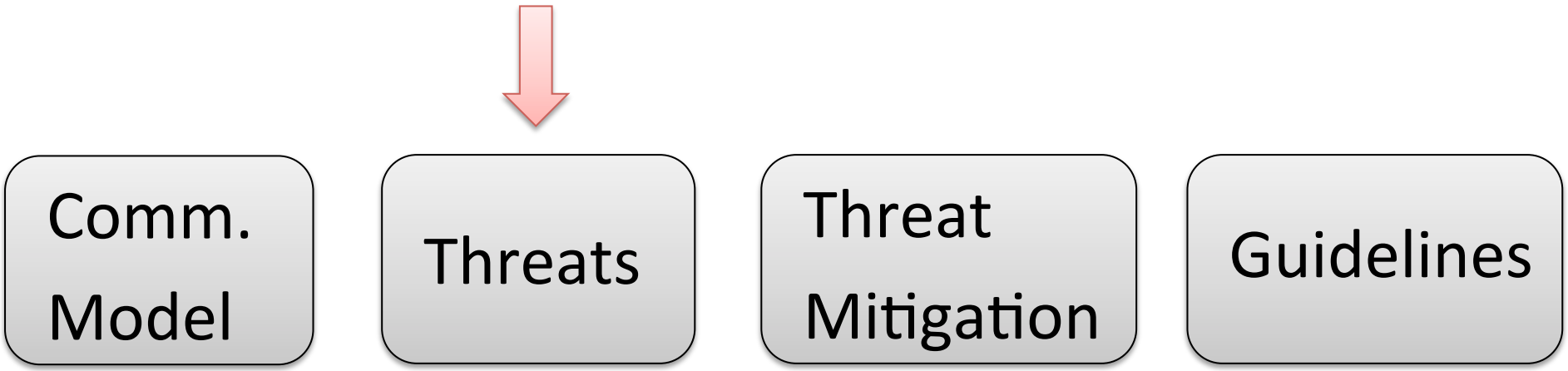
Threats

Threat Mitigation

Guidelines

Communication Models




Comm. Model
Threats
Threat Mitigation
Guidelines

Privacy Threats

- Correlation
- Identification
- Secondary use
- Disclosure
- Exclusion
- Surveillance
- Stored data compromise
- Intrusion
- Misattribution

Should be familiar from security

Privacy Threats (today's sample)

- **Correlation**
- **Identification**
- Secondary use
- Disclosure
- Exclusion
- **Surveillance**
- **Stored data compromise**
- Intrusion
- Misattribution

Correlation

- The combination of various pieces of information related to an individual or that obtain that characteristic when combined.

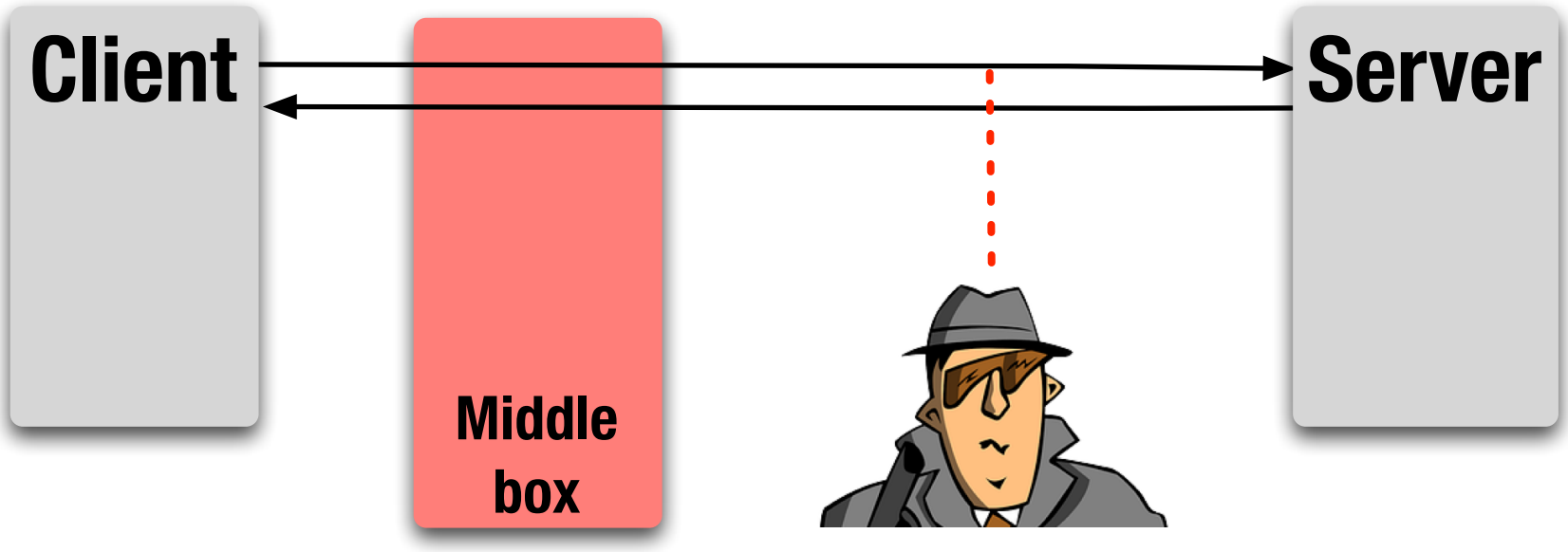
Example: TLS session resumption

Identification

- The linking of information to a particular individual to infer an individual's identity or to allow the inference of an individual's identity.
- Sometimes a threat, sometimes not.
- Lots of protocols support direct identification (e.g., SIP, XMPP) or validation of claims that entities are who they say they are.

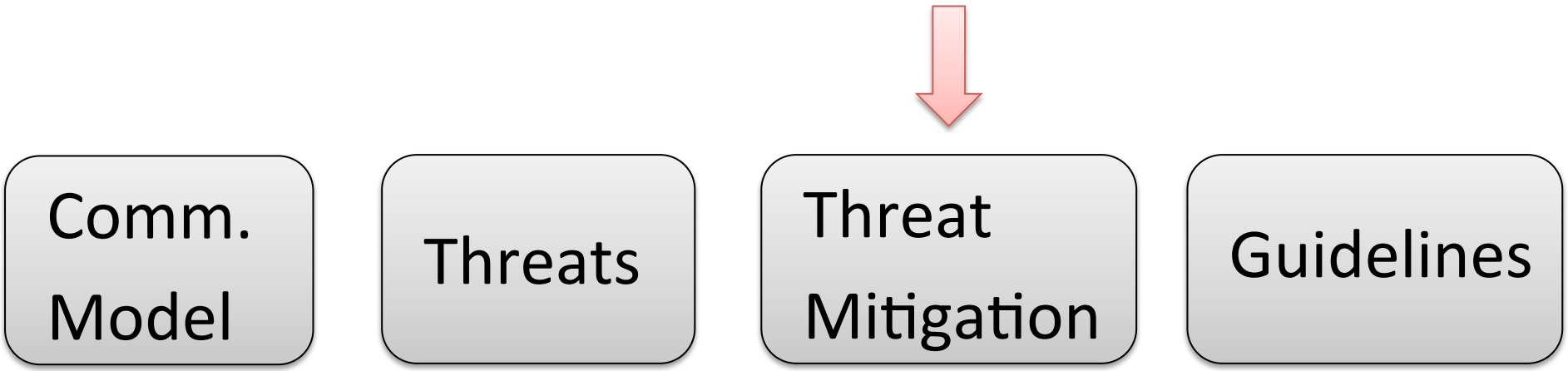
Surveillance

- The observation or monitoring of an individual's communications or activities.
- Includes traffic analysis and observation of encrypted communications.



Stored Data Compromise

- Failure to prevent unauthorized or inappropriate access to stored data.
- Typically outside of IETF scope, but consider key management, access control, operational logging.
  - E.g., RFC 6302 recommends that servers log (forever?) source/dest ports, timestamp, transport protocol in addition to IPv4 address – implications of compromise?


Comm. Model
Threats
Threat Mitigation
Guidelines

Threat Mitigations

- Data minimization
  - See next slide
- Security
  - Confidentiality
  - Authentication
  - Access control
  - Authorization mechanisms
- User participation
  - Control over which personal data is shared
  - Signaling user preferences

Data Minimization

- Collection
  - E.g., why send IP addresses in mail headers?
- Disclosure
  - What data is (unnecessarily) exposed to proxies, relays, other intermediaries?
- Identifiability
  - See next slide
- Sensitivity
  - E.g., send precise geo or a geo region?
- Retention
- Use

Identifiability

- Anonymity: individual cannot be identified within a group
  - Really hard in practice
  - E.g., RFC 3325 SIP ‘From: Anonymous’
- Pseudonymity: individual is identified by some identity-shielding name
  - Very common in Internet protocols
  - Can still yield high identifiability (and facilitate correlation) depending on construction, persistence, use
    - E.g., IPv6 SLAAC using MAC address vs. RFC 4941 temporary address

Identifiability

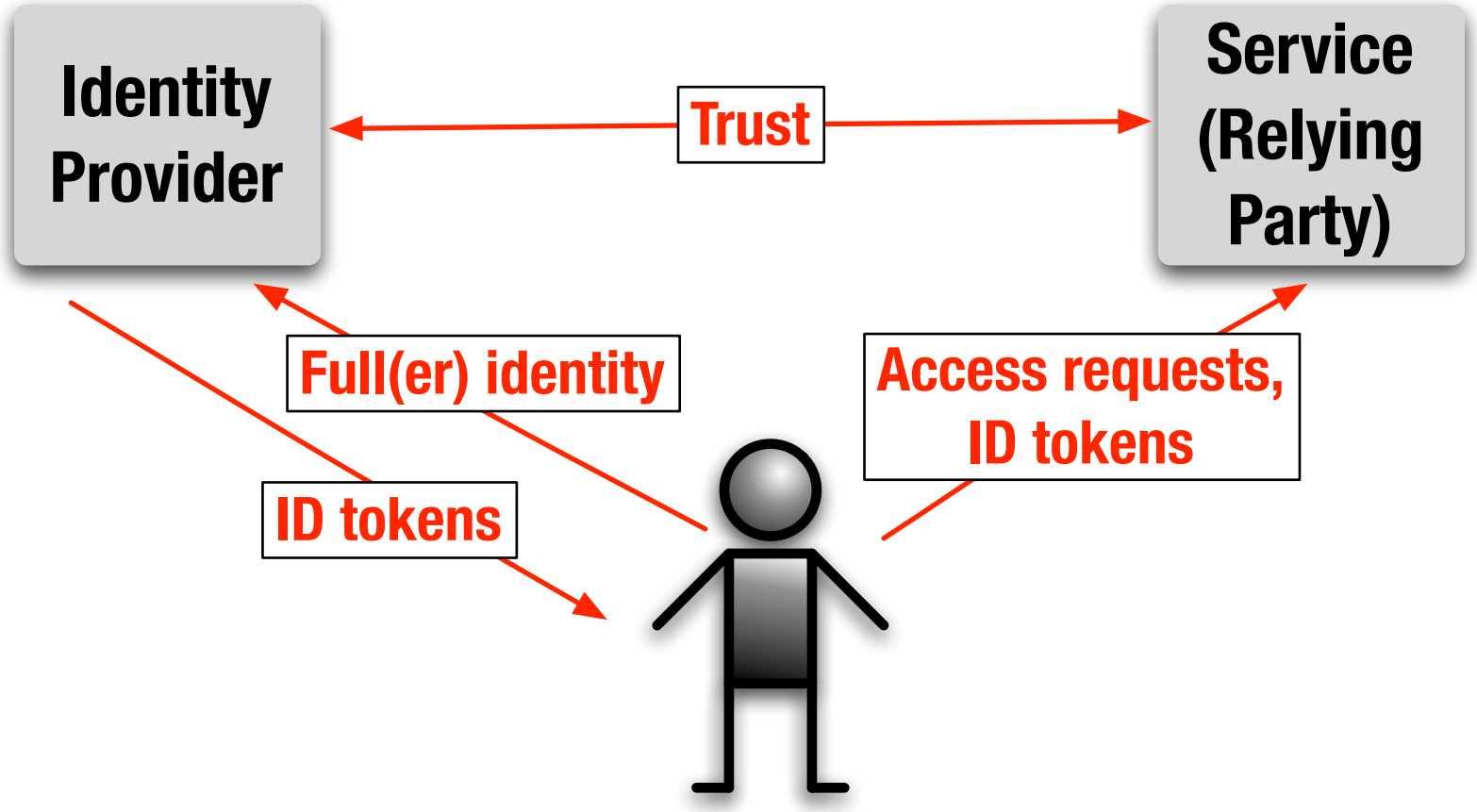
- Identity confidentiality: any party other than the recipient cannot sufficiently identify the sender

TLS 1.3 Example (work in progress).

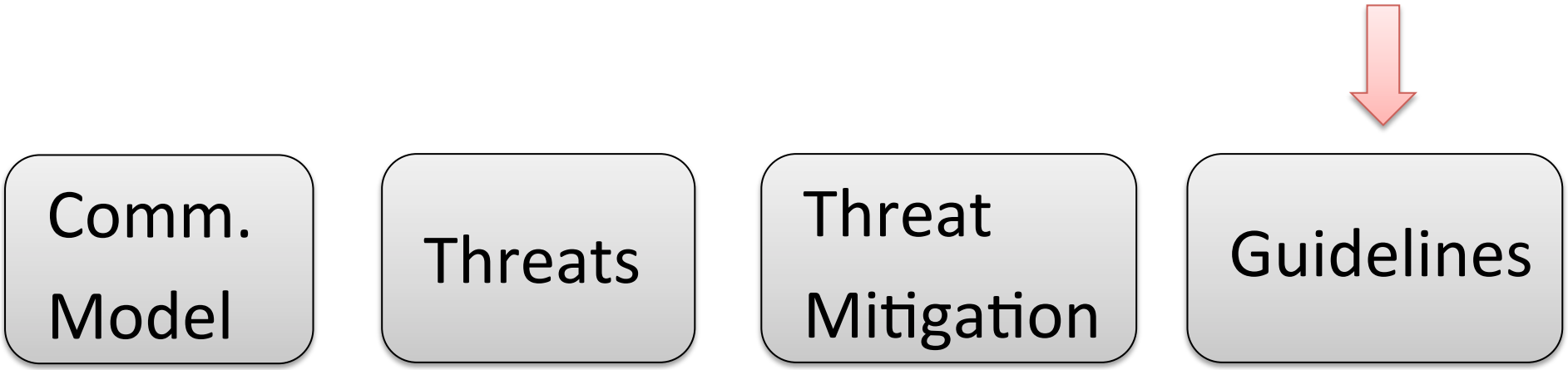
- Further examples: Extensible Authentication Protocol (EAP) and EAP methods.

Identifiability

- Minimization within identity management



- Example: OAuth


Comm. Model
Threats
Threat Mitigation
Guidelines

Guidelines: Questions to ask yourself about protocol being designed

Four categories:

1. Data minimization
2. Security
3. User participation
4. General

Data Minimization Guidelines (sample)

- **Identifiers**
  - Does the protocol use identifiers that allow different protocol interactions to be correlated?
  - What identifiers could be omitted or be made less identifying while still fulfilling the protocol's goals?
- **Persistence of identifiers**
  - Does the protocol allow implementers or users to delete or replace identifiers?
  - How often does the specification recommend replacing identifiers (by default)?
  - Can the identifiers, along with other state information, be set to automatically expire?

Data Minimization Guidelines

- *Identifiers*
- *Persistence of identifiers*
- Data (and personal data)
- Observers – controls on exposure
- Fingerprinting
- Correlation – expected data combinations
- Retention – implications of protocol design

Security Guidelines (sample)

- **Surveillance**
  - Does the protocol leak information that can be observed through traffic analysis, such as packet sizes or timing that allow observers to determine characteristics of the traffic (e.g., which protocol is in use or whether the traffic is part of a real-time flow)?
  - Section 2 of 3552 provides further info.
- **Stored data compromise**
  - How do the protocol's security considerations prevent or mitigate stored data compromise?

Security Guidelines

- *Surveillance*
- *Stored data compromise*
- Intrusion
- Misattribution

User Participation Guidelines

- User control
- Control over sharing with recipients
- Control over sharing with intermediaries
- Preference expression

General Guidelines

- **Trade-offs**
  - Does the protocol make trade-offs between privacy and usability, privacy and efficiency, privacy and implementability, or privacy and other design goals?
- **Defaults**
  - If the protocol can be operated in multiple modes or with multiple configurable options, does the default mode or option minimize the amount, identifiability, and persistence of the data and identifiers exposed by the protocol?
  - Does it provide the strictest security features of all the modes/options?

Resources

- RFC 6973
  - https://tools.ietf.org/html/rfc6973

- Questions, requests for help & reviews:
  - Mail to ietf-privacy@ietf.org