IP Flow Anonymisation Support
draft-boschi-ipfix-anon-03

Elisa Boschi, Brian Trammell – Hitachi Europe, Zürich

[See http://people.ee.ethz.ch/~boschie/draft-boschi-ipfix-anon-03.txt]

Monday 23 March 2009, San Francisco, California, USA

Why?

- Standardization of flow export [RFC5101] and storage representations [draft-ietf-ipfix-file] removes a technical barrier to data sharing for
  - interdomain measurement efforts, and
  - trace publication for research.
- End-user privacy concerns still a barrier.
- Anonymisation provides one toolset for addressing these concerns.
- Need a way to represent the anonymisation status of exported or stored data.

What?

- Node and user identifying information:
  - IP addresses
  - MAC addresses [TODO]
- Additional information that can be used to profile users and/or attack anonymisation techniques:
  - ports (application profiling)
  - timestamps and counters (host profiling, fingerprinting)
  - most other fields (fingerprinting)

How?

- Generalization
  - mapping of multiple real values to a single anonymised value.
  - aggregates; does not preserve countability.
- [One-Way] Substitution
  - one-to-one mapping of real to anonymised values.
  - count of unique values preserved across anonymisation.
- [One-Way] Set Substitution
  - each anonymised value respresents a single real value, but
  - each real value may be represented by multiple anonymised values.
  - does not aggregate, does not preserve countability.

For how long?

- Anonymisation functions map a real space of values to an anonymised space.
- This mapping function may change over time (e.g. through rekeying or deletion of permutation table), affecting long-term comparability of anonymised datasets.
- Stability measures this change.

Anonymisation Record

templateId

informationElementId

informationElementIndex

anonymisationTechnique

anonymisationStability

scope identifies {Template, IE}

index optional (multi-IE only)

Technique used to anonymise

Associated with a Template, exported via IPFIX Options.

Supported Techniques (1)

- None
  - Data assumed to be real at EP
- Undefined
  - no assumption anonymisation at EP
- Deletion/“Black Marker”
  - IPFIX supports this natively – simple removal of IE from Template
- Precision Degradation/Truncation (Generalization)
  - Replacement of host addresses with network addresses
  - Timestamp and counter degradation

Supported Techniques (2)

- Binning (Generalization)
- Enumeration (Set Substitution)
  - Assignment of a unique value to each record.
  - May be applied to preserve ordering, esp. with timestamps.
- Permutation (Substitution)
  - Assignment of a unique anonymised value to each real value.
- Prefixed Permutation (Substitution)
  - Permutation that preserved structure present in source data (e.g. CryptoPAN prefix-preserving IP address anonymisation)

Supported Stability Classes

- Undefined
  - No assumption about stability at EP
- Session
  - All anonymisation parameters stable for duration of Transport Session.
- Exporter-Collector
  - All anonymisation parameters stable for an EP-CP pair.
  - Different sessions between same EP and CP are comparable.
- Stable
  - All anonymisation parameters stable for an EP.
  - Different sessions from same EP are comparable.

Guidelines for IPFIX Anonymisation

- Certain IPFIX data structures can leak information about original data:
  - IPFIX Message Header Export Time
  - IPFIX Message Header Observation Domain, if linked to EP/MP address
  - exportingProcessIPv[46]Address
  - collectionTime and File Time Window Options
- Care should be taken with these to prevent deanonymisation.

Next Steps

- Continued work on Open Issues:
  - MAC address anonymisation
  - Finalization of classification, stability classes, and techniques
  - Editorial completion
    - IANA Considerations, Security Considerations, Examples
    - nonsensical IE/Technique combinations
    - other relevant guidelines
- Consideration as new Working Group item within the context of the Mediator effort, when appropriate.