Personal Information Tagging for Logs

https://tools.ietf.org/html/draft-rao-pitfol-02

Sandeep Rao, Shivan Sahib, Ryan Guest, Santhosh CN

IETF 108, PEARG

Jul 27 2020

Recap

Log and Privacy

- Privacy
  - Improve ways to identify personal information in logs
- Challenges
  - Subjective - many types of log data and formats
  - Dictionary, Heuristics query, Data-Set based training model
  - Vendor specific schema / privacy policy
  - No standard / guidelines - (what) data to be protected and (what) action to enforce

Document Goal

- Enable operators to explicitly detect and filter personal information in logs
- Reference model to enable tagging of personal information at source
- Illustrate ways to attach sensitivity/privacy tagging to log data
- Privacy Control Actions - Consumer Role-based / Case-based

Updates from -01

- Address review comments
- Defining a log privacy schema
- Proposal for a Personal Information Identifier Registry
- Log processing and Access Control

Personal Information Identifier Registry

Example: Define “namespace” for personal information

Name	Abstract Data Type	Description	Sensitivity [1–High 5–Normal]
nationalIdentity	String	National IDs issued by sovereign governments. Eg., SSN	1
drivingLicense	String	Driving License number	1
taxIdentity	String	Tax identification numbers	1
credtCardNumber	String	Credit cards	1
bankAccount	String	Bank account number	1
dateOfBirth	Date	Date of Birth	2
personName	String	Person name	1
emailAddress	String	Email	2
phoneNumber	Number	Phone	1
zipCode	Integer	Zip codes	5
ipAddress	ipv4Address	IPv4 or IPv6 Address	4
dateTimeSeconds	dateTimeSeconds	seconds	5
age	Integer	Age	2
ethnicGroup	String	Ethnic group	1
genderIdentity	String	Gender identity	1
macAddress	macAddress	MAC Address	4

Privacy Tagging

Field level tagging

```
<120> Apr 18 16:32:58 10.0.1.11 QAUDJRN: [AF@0 event="AF-Authority
failure" violation="A-Not authorized to object" actual_type="AF-A"vjrn_seq="1001363"
timestamp="20120418163258988000"vjob_name="QPADEV000B"  {personName="XYZZY"
pii_sensitivity_level=1} {emailAddress="xyz@foo.com" pii_sensitivity_level=2]
object_type="*FILE" pgm_name="" pgm_libr="" workstation=""]
```

Log level tagging

```
<120> Apr 18 16:32:58 10.0.1.11 QAUDJRN: [AF@0 event="AF-Authority
failure" violation="A-Not authorized to object" actual_type="AF-A"vjrn_seq="1001363"
timestamp="20120418163258988000"vjob_name="QPADEV000B"  personName="XYZZY"
emailAddress="xyz@foo.com" object_type="*FILE" pgm_name="" pgm_libr="" workstation=""],
{pii_sensitivity_level=1}
```

Privacy Tagging

Redaction Action

```
<120> Apr 18 16:32:58 10.0.1.11 QAUDJRN: [AF@0 event="User Logged In”
    timestamp="20200418163258938000"
    personname="xyz@foo.com" ip_addr=”10.21.23.1” phonenum=”+912232422213”],
pii_metadata= {
   "piiDescriptor": {
         "field": "username",
    },
    "action": {
             "SensitivityLevelTag": "4",
             "Action": "FullAnonymization",
    }
}
```

Future Work

- Privacy preservation across log transformations
- Change of Privacy marking policy / classifications
- Out-of-band mechanism to notify privacy schema

Open Questions

- Continue to pursue this work in PEARG?
- Does the scope need to change?
- Request for RG adoption ?