3GPP TSG-SA WG3 LI Meeting #75 **s3i190623**

October 29 – November 1, 2019; West Palm Beach, FL, USA

Source: **Public Safety Canada**

Title: **Content Filtering and Reduction**

Document for: **Discussion**

Agenda Item: **11.2**

Work Item / Release: **LI16**

Abstract of the contribution: *This contribution is a discussion on the need and challenges with filtering out content to avoid over collection, and reduce excessive Call Content*

LI Content Filtering and Exclusion

Introduction

There are ongoing and growing concerns with the Lawful Interception and delivery of broad / all inclusive IP data. This discussion paper is attempting to capture the work / discussions to date to progress the discussion to concrete stage 2 & 3 work both in 3GPP SA3-LI and ETSI TC-LI.

First issue is with “over collection” as interception of the whole IP data “pipe” may include IP based services which are outside the scope of the warrant. E.g., a data only intercept should not include CSP provided Voice (VoIP) aspects. Same applies for CSP provided Location services & data. Depending on National law and case law this over collection shall not happen. This applies to both for IRI and CC.

From 33.126:

R6.4 – 10 LI Service Scope - The CSP shall only deliver Interception Product relating to specific CSP services which are specified implicitly or explicitly in the warrant.

R6.4 - 130 Separate delivery of services - The CSP shall be able to support delivering Interception Product for a particular service separately from other services' Interception Product (e.g. delivering SMS Interception Product independent of CS Voice Interception Product).

This last requirement brings up the need for the CSP to also be service (Data, Voice, Messaging, etc.) aware at the IP level for independent delivery.

Second issue is the ongoing discussions on how to deal with the target IP data ‘explosion’ / exponential growth both within the CSP but also within the LEA (and in the hand-over interfaces in-between) .

Again from 33.126

R6.2 - 180 Completeness - The CSP shall be able to intercept all Target Communications as specified in the warrant.

A.4 Delivery (from informative annex A)

"A mechanism to reduce Interception Product volume based on information received from the LEA (e.g. exclusion of particular flows such as movies) may be bilaterally agreed."

Note 1: There is a 'sensitivity' with using "filtering" to describe the reduction of flows as it might raise questions about what and how content was reduced; i.e., is the content over filtered. "Exclusion" and "semantically-aware compression" have been proposed.

Note 2: In some (many ?) National jurisdictions the CSP may be compelled to deliver all content to the LEA. This raises the question on how do this most efficiently.

The first and second are independent issues but may have a common architectural approach to address.

Details, issue 1

With filtering to avoid over collection there is a set of issues / challenges:

Note 3: in this context using "filtering" should OK, given it is to meet the warrant specified requirements.

1. Agreement on the set of services that can be tasked and or filtered. Currently 33.128 defines four services (Data, Voice, Messaging, and PTT). Is this sufficient? Or do we need a more extensible mechanism to communicate and expand the services scope and the resulting the filtering implementation?

2. What actually defines a unique service such that it can be discovered and / or filtered at the IP data flow level?

3. How is the set of collection rules communicated to the appropriate filtering engine. I ASSUME it's via X1 as part of the tasking information as to which services to intercept and the rest will be filter out.

4. Where is the best place to execute said filtering engine? The two obvious places are the POI and / or MDF. At the POI implies smarter, more complex execution engines (not just mirror and forward). At the MDF a (slightly) more centralised and hopefully scalable approach.

Note 4. Lingering concern that doing filtering southbound of the POI may still violate the warrant as the CSP has collected the data.

A solution to this overall issue is really packet level inspection with a policy to say collect and hand-over or not. And except for messaging, should (perhaps maybe) need not to inspect every packet once a flow has been identified. For multiple warrants against the same target, but with different service scopes, the X1 tasking info / profile for each task should be supported and independent. The industry has been doing this type of packet inspection and filtering for years. Examples (thanks Rogers) and there many more (BEA etc.)

> The **sigrok** project aims at creating a **portable, cross-platform, Free/Libre/Open-Source signal analysis software suite** that supports various device types (e.g. logic analyzers, oscilloscopes, and many more).
>
> It is licensed under the terms of the **GNU GPL, version 3 or later**. Design goals and features include:
>
> - **Broad hardware support**. Supports many different devices (logic analyzers, oscilloscopes, multimeters, data loggers etc.) from various vendors.
> - **Cross-platform**. Works on Linux, Mac OS X, Windows, FreeBSD, OpenBSD, NetBSD, Android (and on x86, ARM, Sparc, PowerPC, ...).

- **Scriptable protocol decoding**. Extendable with stackable protocol decoders written in Python 3.
- **File format support**. Supports various input/output file formats (binary, ASCII, hex, CSV, gnuplot, VCD, WAV, ...).
- **Reusable libraries**. Consists of the libsigrok and libsigrokdecode shared libraries which can be used by various frontends/GUIs.
- **Various frontends**. PulseView (LA GUI), sigrok-meter (DMM GUI), sigrok-cli (command-line), and other frontends all build upon the above libraries.

https://sigrok.org/wiki/Main_Page

https://sigrok.org/wiki/Protocol_decoders

L7-filter is a classifier for Linux's Netfilter that identifies packets based on application layer data. It can classify packets as Kazaa, HTTP, Jabber, Citrix, Bittorrent, FTP, Gnucleus, eDonkey2000, etc., regardless of port. It complements existing classifiers that match on IP address, port numbers and so on.

Details issue 2

For the exponential (not really useful traffic (e.g., publicly available)) handling based on National requirements. This is in regards to CC only the IRI (metadata) should be delivered in all cases.

The first approach is to deliver everything. This requires a means to do this efficiently with in the the CSP, hand-over interface (HI-3) and within the LEA itself. ETSI has a study item to explore this: (nothing published yet)

TR 103 656

Title	Lawful Interception (LI); Study on high bandwidth delivery
Scope and Field of Application	The scope of the study item is to assess and characterise the problems associated with interception and secure onward delivery of high-bandwidth user traffic using TCP or TLS as currently defined in ETSI TC LI specifications, identify whether there is a need to solve these problems, and identify potential technical and other measures that can be used to mitigate or address them.
Supporting Organizations	Ericsson LM, BAE Systems AI Ltd, Telefónica S.A., Vodafone Group Plc, Deutsche Telekom AG, PIDS, OTD, Utimaco TS GmbH, National Technical Assistance, trovicor, BKA, BT plc, RCS S.p.A, Ministère Economie et Finances, Security Service (Sweden), Polisen (Swed

One of the open issues that current HI-3 interface, being TLS based, has through-put scalability issues where as using UDP has guaranteed delivery concerns. The ETSI TC-LI work on supporting OTTs may play into this solution.

The second approach is to use some form of exclusion / compression / reduction approach within the CSP to not deliver said traffic at least over the hand-over (HI-3) i/f and maybe not across the X3 i/f.

This has many of the same issues / concerns as with content filtering, but with added challenges.

1. Agreement on the set of content delivery that can be excluded. Do we need an extensible mechanism to communicate and expand the exclusion rules?

2. What actually defines a unique content flow / service such that it can be discovered and excluded from the IP data flow level?

3. How is the set of collection rules communicated to the appropriate exclusion engine. I ASSUME it's via X1 as part of the tasking information as some form of template or maybe even a script, but there may be other options. (see 2 above.)

4. Where is the best place to execute said exclusion engine? The two obvious places are the POI and / or MDF. At the POI means implies smarter, more complex execution engines (not just mirror and forward). At the MDF a (slightly) more centralised and hopefully scalable approach, but implies all this traffic traverses the CSP core (again).

A solution to this issue is really packet level inspection with a policy to say collect and hand-over IRI but not CC for specified services

Please see the Annex for two OTD discussion papers on the need for this approach.

Note: often the CSP is a conduit for delivering the streamed data as it is really delivered by a Content Delivery Network (CDN) but is sourced from a Content Provider. What does the meta-data really represent? Also, content providers and CDNs do not always maintain past content; e.g. Netflix monthly changes it's "line-up" and YouTube takes down content all the time because of Digital Rights abuse requests. So the meta data may be stale quickly.

A third (of many) options is to hand over the meta-data (IRI) but cache the content within the CSP (some-place) in case the LEA wants to retrieve it later (CSP pushes the IRI, LEA pulls the CC if/when required). This approach brings up many questions, such as:

- Where is it cached ?
- Who pays for the cache ?
- How big a cache ?
- Mechanism for LEA to access / request cache content ?
- Are the Content Providers 'happy' with their data being cached ?
- Policy for removal from cache ?
- Etc.

Discussion

For the first issue the potential LI solution level approach appears to be determined. As per 33.128 5.2.4.2 (below) the LEA requests the service(s) to be intercepted, and via X1 the POIs/TFs and / or MDFs are tasked with collecting the specified services. All other content should be filtered out (though it doesn't explicitly say that (yet)).

5.2.4.2 CSP service type

- Voice.
- Data.
- Messaging (e.g. SMS/MMS).
- Push-to-Talk (including MCPTT).

The LIPF shall be able to provision the POI, TFs and the MDF2/MDF3 with the service type as applicable to a warrant.

When multiple service types are applicable to a target due to multiple warrants, if those multiple service types are provisioned to the POI and TFs, then the MDF2/MDF3 shall be able to restrict the delivery of IRI/CC based on the service type.

This looks workable, though the detailed definition of the service is still an issue. Are they standardised ? Should they be ?

For the second issue (content exclusion) it is trickier. I doubt that the list of content streams / streamers could be standardised nor maintained. 33.126 (annex A) states that what should be excluded is via agreement between the LEA and CSP. But how does that work ?

One option is that the LEA defines the exclusion rules to be delivered to the CSP (e.g., via HI-1), these rules are included in the tasking as an object sent to the POIs/TFs and / or MDFs. An exclusion engine interprets the object and either supresses the delivery of said content or sends it to a cache (somewhere). (If caching, then the cache id and content id will need to be communicated to the LEA.)

Note that the proposed Aggregation Function may be good place to host the Execution Engine. (No I didn't get payed to say that !)

This approach has little impact the to the LI interfaces but still has a big issue with how to define what traffic to exclude. Is it sufficient to specify NetFlix, Apple-TV, BBC, SportsNet, etc. and let the Exclusion Engine figure it out ? Or should more details of how these are actually identifiable be specified. Also do we need more granularity ?

Summary

For both issues there are solutions but the devil is in the implementation details.

The first issue is really only collect and report this (services IRI and CC). The second issue is really don't report this content. Both require some form of flow inspection and subsequent policy decision on delivery or not.

Both require agreed definitions of what to collect & report and what not to. Both require a view as to where best to filter/exclude.

Conclusions

Issue one appears well in hand, just every-one needs to agree.

Issue two needs more work to agree on the mechanism(s) and how to define the content to be excluded. The mechanism work will require ESTI support.

Annex

LI(19)R46027
Exclusion of Flows.doc

LI(19)P52046_Report
ing_or_excluding_an_i