

Models for adaptive-streaming-aware CDNI - Logging

draft-brandenburg-cdni-has-01, section 3.4

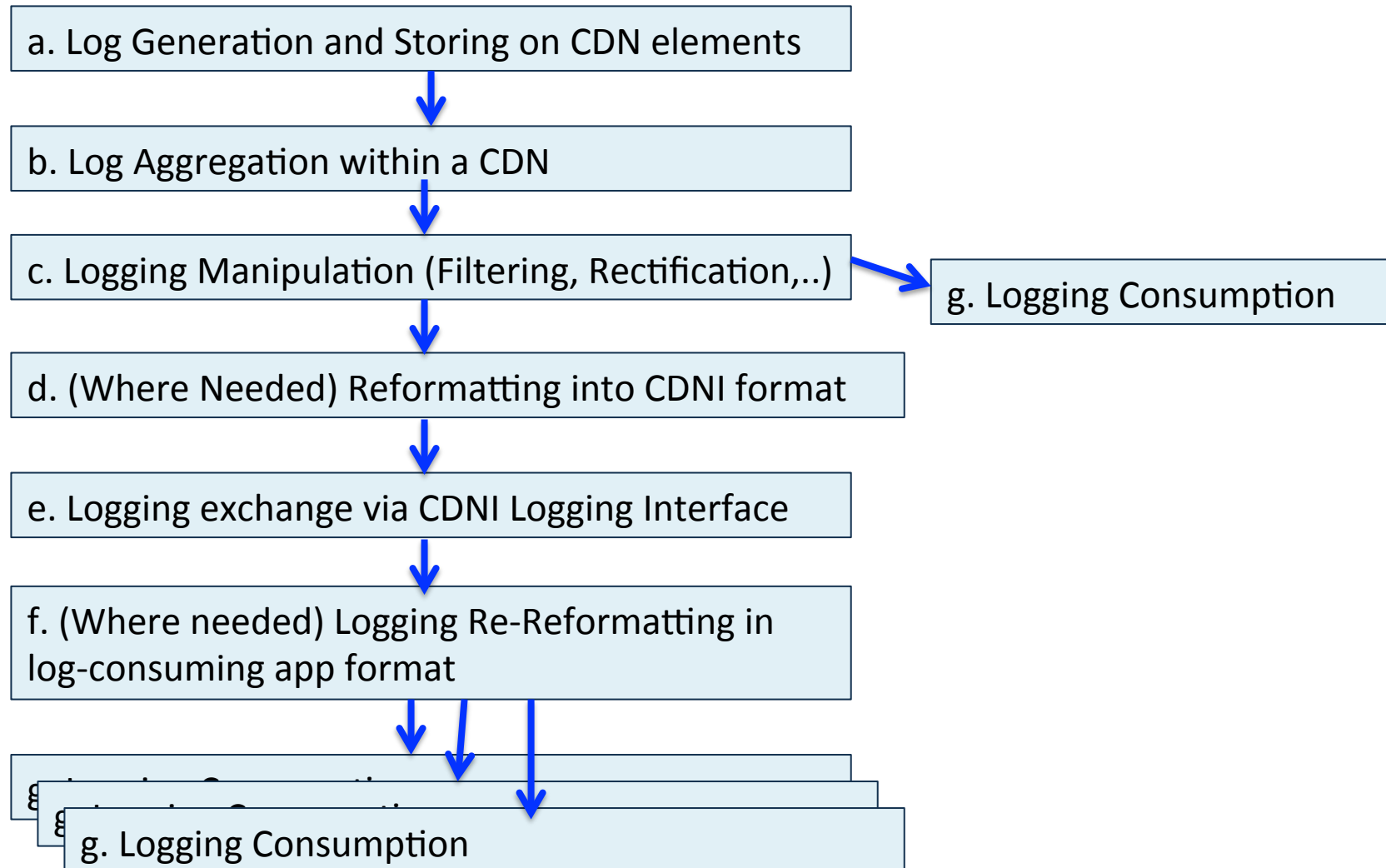
CDNI Extended Design Team Meeting
Virtual Meeting
May 29, 2012

Francois Le Faucheur (flefauch@cisco.com)

Key Consideration re HAS vs CDNI

- The key consideration for HAS with respect to logging is the **potential increase of the number of Log records by 2 to 3 orders of magnitude**, as compared to regular HTTP delivery of a video, since log records would typically be generated on a per-chunk-delivery basis instead of per-content-item-delivery basis.
- This impacts the scale of every processing step in the Logging Process

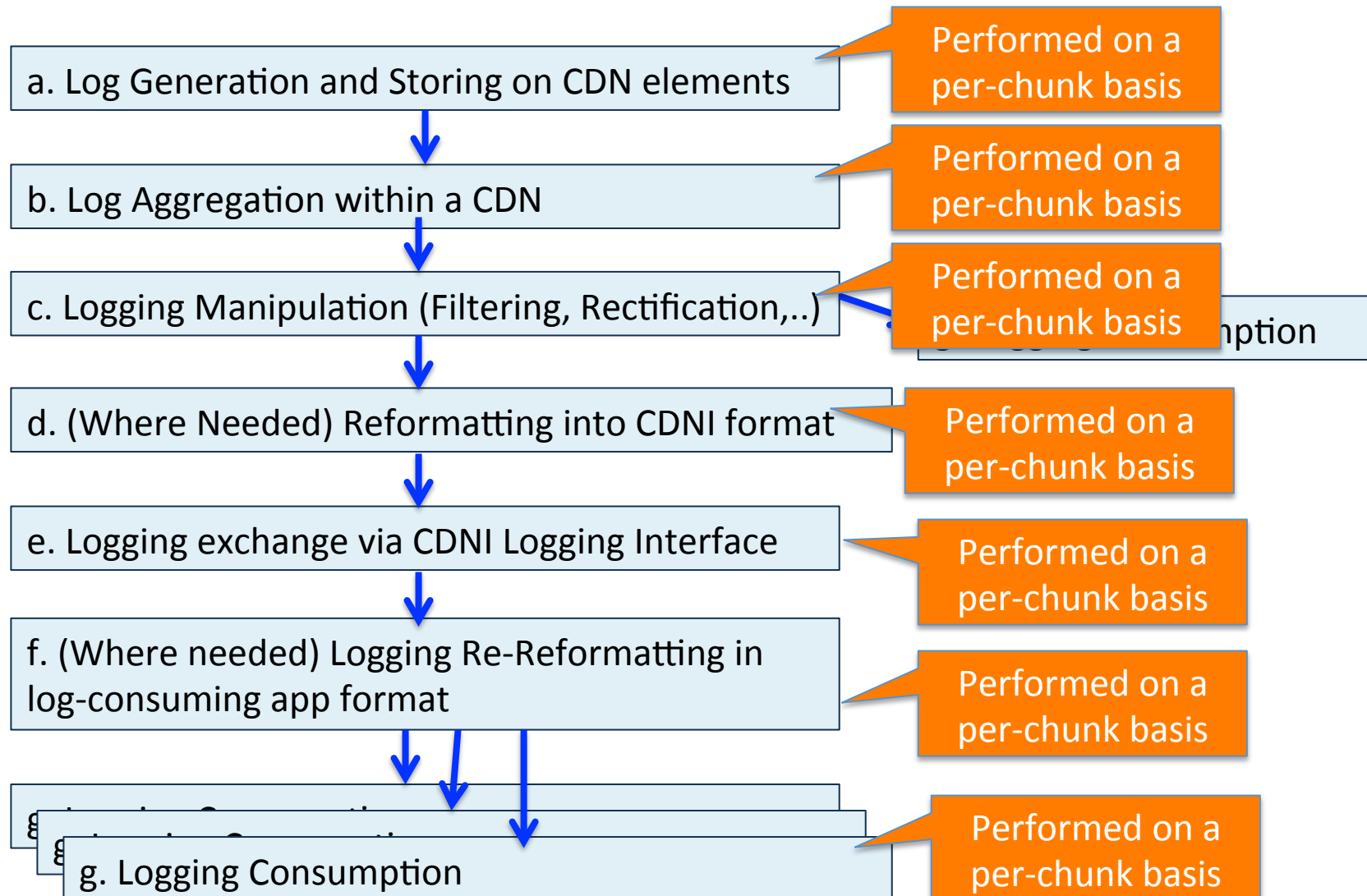
Logging Process



Option 4.1: “Do-Nothing” Approach

- In this approach, each HAS-chunk delivery is considered, for CDNI Logging, as a standalone content delivery.
- In particular, a separate log record for each HAS-chunk delivery is included in the CDNI Logging Interface in step [e].

Option 4.1: “Do-Nothing” Approach



Option 4.1: “Do-Nothing” Approach

Effect on CDNI Interfaces: None

Effect on uCDN and dCDN: None

Advantages/Drawbacks:

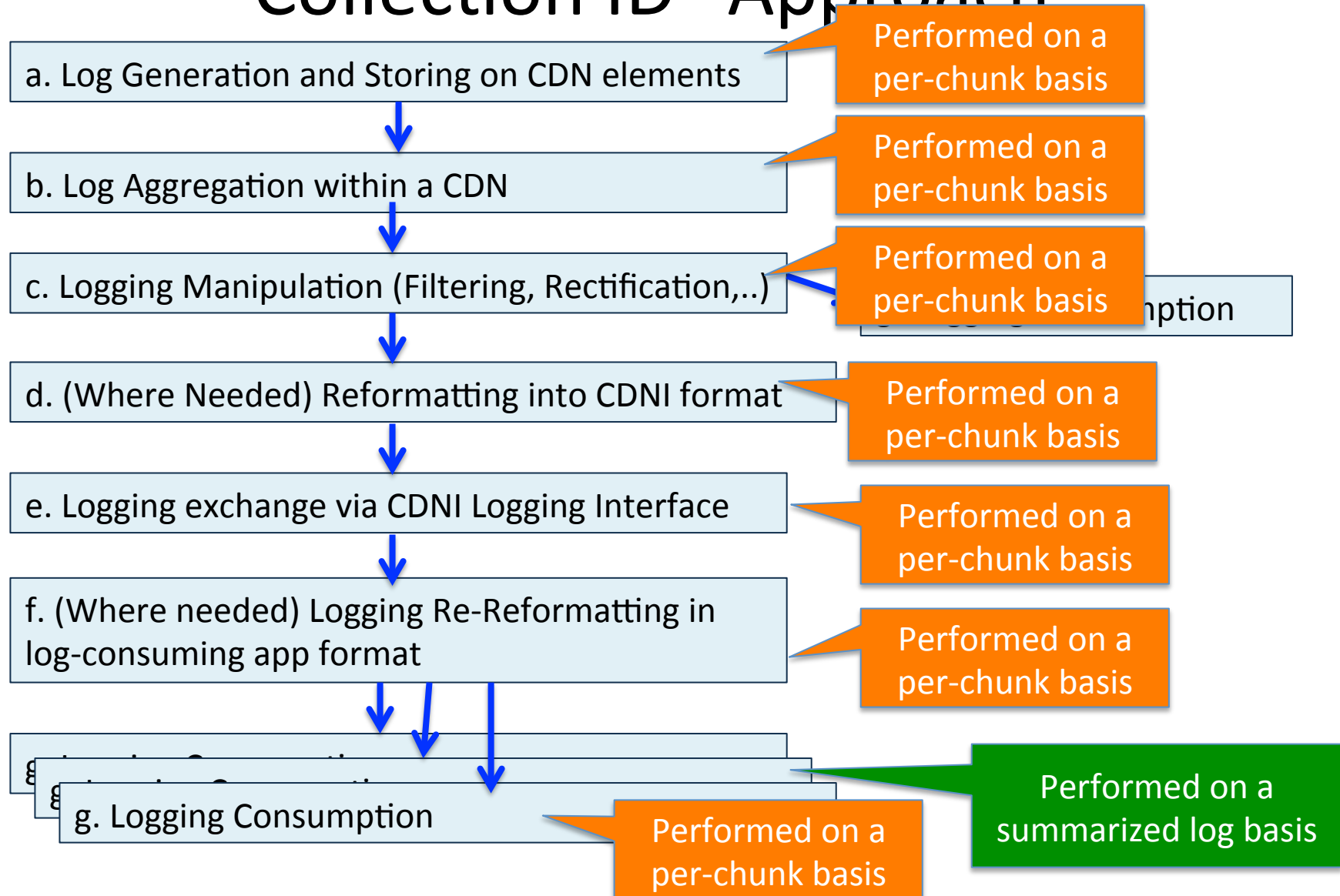
- + No information loss (i.e. all details of each individual chunk delivery are preserved). While this full level of detail may not be needed for some Log consuming applications (e.g. billing), this full level of detail is likely valuable (possibly required) for some Log consuming applications (e.g. debugging)
- + Easier integration (at least in the short term) into existing Logging tools since those are all capable of handling per-chunk records
- + No extension needed on CDNI interfaces
- High volume of logging information to be handled (storing & processing) at every step of the Logging process from [a] to [g] (while summarization in step [f] is conceivable, it may be difficult to achieve in practice without any hints for correlation in the log records).

While the high volume of logging information is a potential concern, we are seeking expert input on whether it is a real practical issue, and if yes, then in what timeframe/assumptions.

Option 4.2: "CDNI Metadata Content Collection ID" Approach

- In this approach, a "Content Collection ID" (CCID) field is distributed through the CDNI Metadata Interface and the same CCID value is associated with every chunk of the same Content Collection.
- The objective of this field is to facilitate summarization of per-chunk records at step [f] into something along the lines of per-HAS-session logs, at least for the Log Consuming application that do not require per-chunk detailed information (for example billing).

Option 4.2: "CDNI Metadata Content Collection ID" Approach



Option 4.2: "CDNI Metadata Content Collection ID" Approach

Effect on CDNI Interfaces: 1 additional metadata field (CCID) in CDNI Metadata Interface

Effect on uCDN and dCDN: None

Advantages/Drawbacks:

- + No information loss (i.e. all details of each individual chunk delivery are preserved). While this full level of detail may not be needed for some Log consuming applications (e.g. billing), this full level of detail is likely valuable (possibly required) for some Log consuming applications (e.g. debugging)
- + Easier integration (at least in the short term) into existing Logging tools since those are all capable of handling per-chunk records
- + Very minor extension to CDNI interfaces needed
- + Facilitated summarization of records related to a HAS session in step [f] and therefore ability to operate on lower volume of logging information in step [g] by log consuming applications that do not need per-chunk record details (e.g. billing)
- High volume of logging information to be handled (storing & processing) at every step of the Logging process from [a] to [f]

Option 4.2: "CDNI Metadata Content Collection ID" Approach

Effect on CDNI Interfaces: 1 additional metadata field (CCID) in CDNI Metadata Interface

Effect on uCDN and dCDN: None

Advantages/Drawbacks:

- + No information loss (i.e. all details of each individual chunk delivery are preserved). While this full level of detail may not be needed for some Log consuming applications (e.g. billing), this full level of detail is likely valuable (possibly required) for some Log consuming applications (e.g. debugging)
- + Easier integration (at least in the short term) into existing systems as those are all capable of handling per-chunk records
- + Very minor extension to CDNI interfaces needed
- + Facilitated summarization of records related to a HAS session in step [f] and therefore ability to operate on lower volume of logging information in step [g] by log consuming applications that do not need per-chunk record details (e.g. billing)
- High volume of logging information to be handled (storing & processing) at every step of the Logging process from [a] to [f]

How much does it help?
What info is needed?

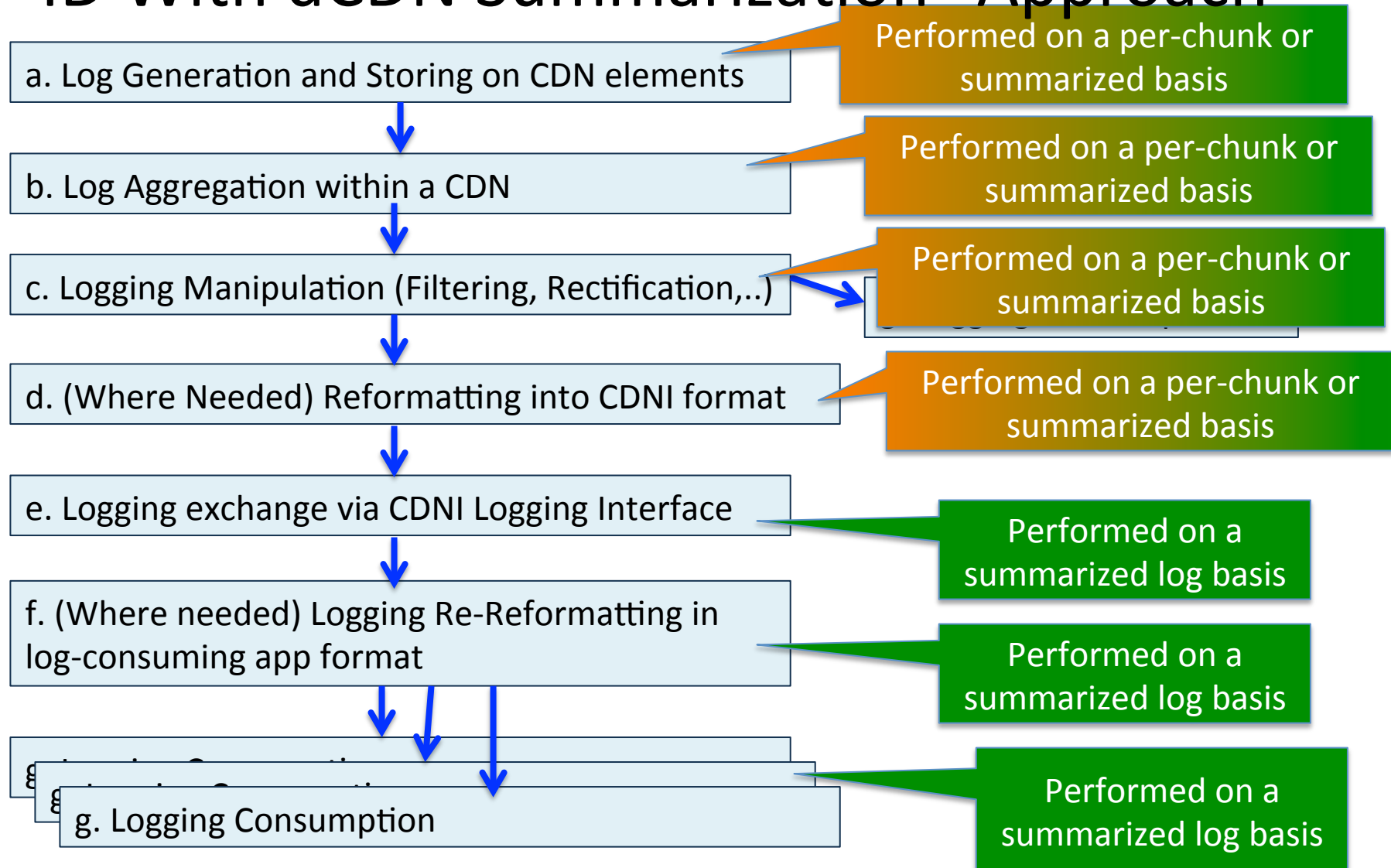
Option 4.3 "CDNI Metadata Content Collection ID With dCDN Summarization" Approach

- In this approach, a "Content Collection ID" (CCID) field is distributed through the CDNI Metadata Interface and the same CCID value is associated with every chunk of the same Content Collection.
- A summarization of per-chunk records is performed at step [d] (or in earlier steps) taking advantage of the CCID, so that a reduced volume of logging information is to be handled in steps [e] to [g] of the logging process (and is optionally also possible in steps [a] to [c]).
- The objective of this approach is to reduce the volume of logging information early in the Logging process

Option 4.3 "CDNI Metadata Content Collection ID With dCDN Summarization" Approach

- Regarding the summarization performed at step [d] (or in earlier steps), there is a **continuum in terms of trade-off between level of summarization of per-chunk records and information loss**.
- For example, it appears possible to perform a summarization that results in significant gains with limited information loss, perhaps using summarized logs along the lines of the Event-Based Logging format discussed in section 3.2.2 of [I-D.draft-lefaucheur-cdni-logging-delivery].
- Alternatively, it may be possible to perform a summarization that results in very significant gains with significant information loss, perhaps using summarized logs along the lines of the Summary-Based Logging format discussed in section 3.2.3 of [I-D.draft-lefaucheur-cdni-logging-delivery].

Option 4.3 "CDNI Metadata Content Collection ID With dCDN Summarization" Approach



Option 4.3 "CDNI Metadata Content Collection ID With dCDN Summarization" Approach

Effect on CDNI Interfaces:

- 1 additional metadata field (CCID) in CDNI Metadata Interface
- Summarized logging information in CDNI Logging Information

Effect on uCDN and dCDN: None

Advantages/Drawbacks:

- + Lower volume of logging information to be handled (storing & processing) at every step of the Logging process from [e] to [g], and optionally from [a] to [d] also Easier integration (at least in the short term) into existing Logging tools since those are all capable of handling per-chunk records
- + Small extensions to CDNI interfaces needed
- Some information loss. The actual information loss depends on the summarization approach selected. While full level of detail may not be needed for some Log consuming applications (e.g. billing), the full level of detail is likely valuable (possibly required) for some Log consuming applications (e.g. debugging)
- Less easy integration (at least in the short term) into existing Logging tools since those are all capable of handling per-chunk records and may not be capable of handling CDNI summarized records

Option 4.3 "CDNI Metadata Content Collection ID With dCDN Summarization" Approach

Effect on CDNI Interfaces:

- 1 additional metadata field (CCID) in CDNI Metadata Interface
- Summarized logging information in CDNI Logging Information

Effect on uCDN and dCDN: None

Advantages/Drawbacks:

- + Lower volume of logging information to be handled (starting from every step of the Logging process from [e] to [g])
also Easier integration (at least in the short term) to e since those are all capable of handling chunk records
- + Small extensions to CDNI interfaces needed
- Some information loss. The actual information loss depends on the summarization approach selected. While full level of detail may not be needed for some Log consuming applications (e.g. billing), the full level of detail is likely valuable (possibly required) for some Log consuming applications (e.g. debugging)

Is this acceptable (eg for debugging)?
Or do we need two mechanisms to pass Logging info
(summarized push, Detailed Pull)?

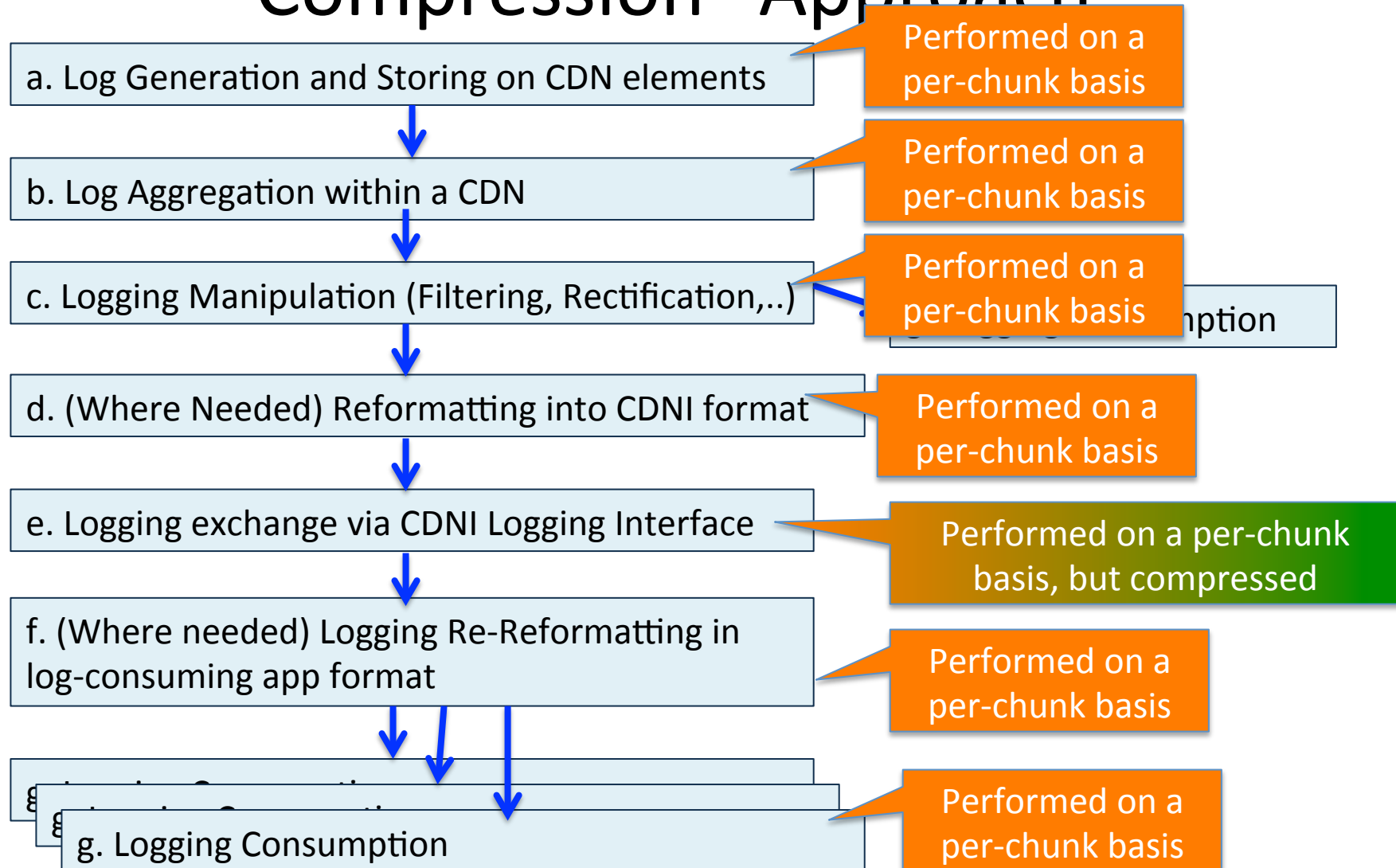
Input is sought on trade-off information loss vs per-HAS-session summarization gains

those are all capable of handling per-chunk records and may not be capable of handling CDNI summarized records

Option 4.4 "CDNI Logging Interface Compression" Approach

- In this approach, a loss-less compression technique is applied to the sets of Logging records (e.g. Logging files) for transfer on the IETF CDNI Logging Interface. The objective of this approach is to reduce the volume of information to be stored and transferred in step [e].

Option 4.4 "CDNI Logging Interface Compression" Approach



Option 4.4 "CDNI Logging Interface Compression" Approach

Effect on CDN Interfaces: 1 additional compression mechanism to be included in the CDNI Logging Interface

Effect on uCDN and dCDN: None

Advantages/Drawbacks:

- + No information loss (i.e. all details of each individual chunk delivery are preserved). While this full level of detail may not be needed for some Log consuming applications (e.g. billing), this full level of detail is likely valuable (possibly required) for some Log consuming applications (e.g. debugging)
- + Easier integration (at least in the short term) into existing Logging tools since those are all capable of handling per-chunk records
- + Small extension needed on CDNI interfaces
- + Reduced volume of logging information in step [e]
- High volume of logging information to be handled (storing & processing) at every step of the Logging process from [a] to [g] except [e].

Input is sought on expected compression gains achievable in practice over sets of logs containing per-chunk records..

Option 4.5 "Full HAS awareness"

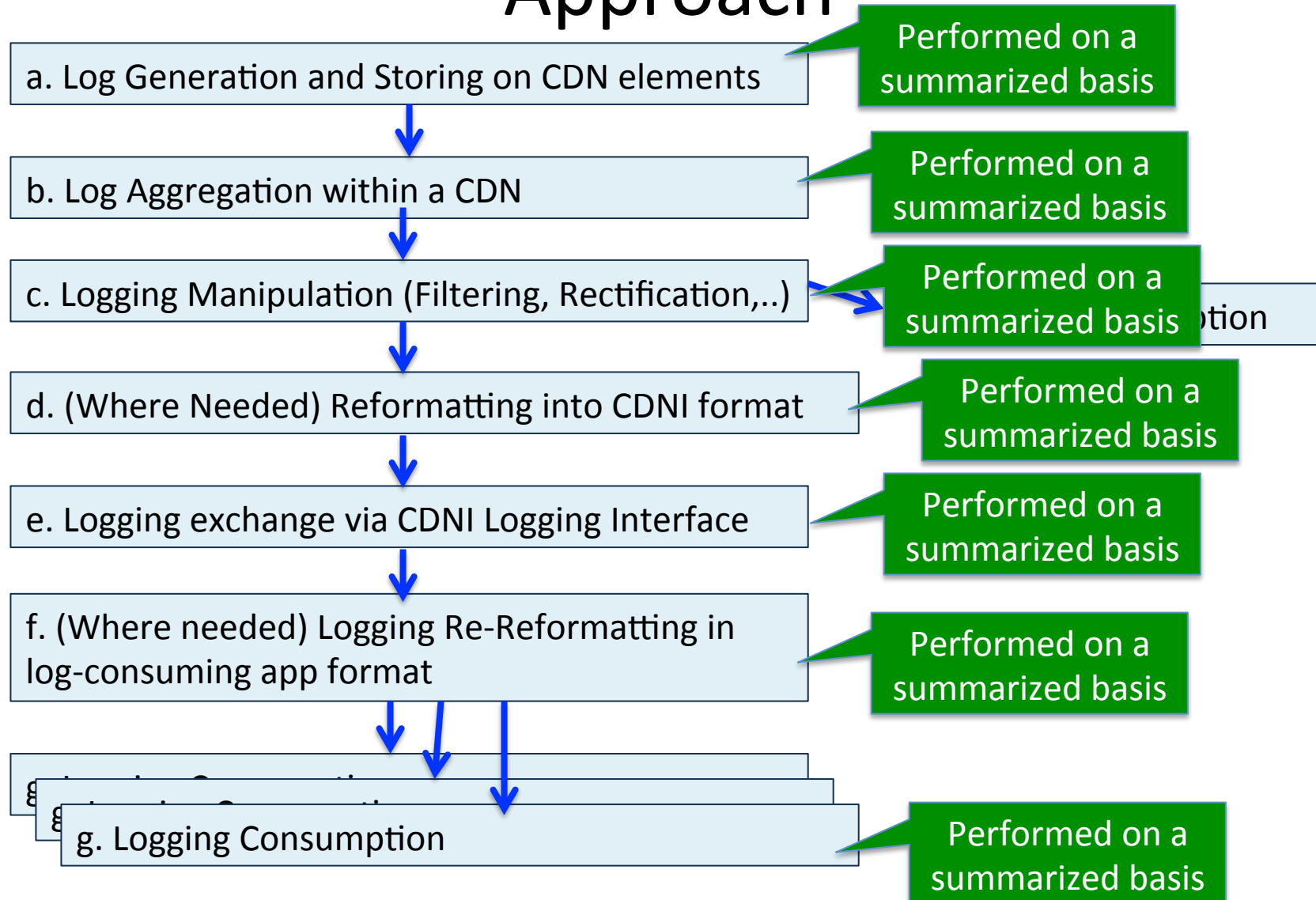
Approach

- In this approach, HAS-awareness is assumed across the CDNs interconnected via CDNI and the necessary information to describe the HAS relationship across all chunks of the same Content Collection is distributed through the CDNI Metadata Interface. In this approach, the dCDN Surrogates leverage the HAS information distributed through the CDNI metadata and their HAS-awareness to generate summarized logging information in the very first place. The objective of that approach is to operate on lower volume of logging information right from the very first step of the Logging process.

Option 4.5 "Full HAS awareness" Approach

- The summarized HAS logs generated by the Surrogates in this approach are similar to those discussed in the section "CDNI Metadata Content Collection ID With dCDN Summarization" Approach" and the same trade-offs between information loss and summarization gain apply.

Option 4.5 "Full HAS awareness" Approach



Option 4.5 "Full HAS awareness"

Approach

Effect on CDNI Interfaces: extend CDNI Logging Interface to support Summarized logs for HAS

Effect on uCDN and dCDN: dCDN is HAS-aware

Advantages/Drawbacks:

- + Lower volume of logging information to be handled (storing & processing) at every step of the Logging process from [a] to [g]
- + Accurate generation of summarized logs because of HAS awareness on Surrogate
- Very significant extensions to CDNI interfaces needed including per HAS-protocol specific support
- Very significant additional requirement for HAS awareness on dCDN
- Some information loss (i.e. all details of each individual chunk delivery are not preserved). The actual information loss depends on the summarization approach selected
- Less easy integration (at least in the short term) into existing Logging tool

Input is sought on trade-off between information loss vs per-HAS-session summarization gains

Points not yet reflected in cdni-has

- In Paris, there seemed to be agreement for CDNI to support flexible sets of Log record entries to be requested by uCDN. This allows CDNI logs to only include necessary info, so to some extent, this helps scaling the “Do-Nothing” Approach a bit.
- There are challenges in generating per-HAS-session logs (even assuming full HAS awareness) such as when the session is load-balanced across surrogates of a CDN or across CDNs [Roy]. This affects Options 4.2, 4.3 and 4.5

Key Questions?

- While the high volume of logging information is a potential concern, is it a real practical issue?
if yes, then in what timeframe/assumptions?
- What are compression gains achievable in practice over sets of logs containing per-chunk records?
- Is there an aggregation/summarization sweet-spot in terms of log summarization vs Information-loss?

Key Questions?

- Do we need 1 mechanism or two mechanisms to pass Logging info (summarized push, Detailed Pull)?
- [Roy]:

“The discussion of possibly allowing the uCDN to specify (via metadata transfer or other means) which log fields or formats it wishes to receive actually, I feel, calls for stepping back and reviewing whether we should have purely a logging interface or more of a reporting or query interface. That would, I imagine, allow the uCDN to specify:

 - 1) what transactions do I want data about (specified by date range and possibly other filter criteria)
 - 2) what attributes and measures do I want to receive in my data
 - 3) what level of aggregation do I want the data at (including for example, at the session level or individual file level)

“

Recommendations

- The “Do Nothing” Approach is a natural trivial starting point that “will do the job” and is MANDATORY
- If trivial “Content Collection ID” or simple compression helps in optional summarization/compression, then could be easily included
- Let’s get real evidence of a likely practical short-term scaling issue with “Do Nothing” Approach, before we add significant extensions. If supported, an additional approach would most likely be OPTIONAL (ie may or may not be supported by a CDN).