Savant: A Framework for Supporting Content Accountability in Information Centric Networks

Diarmuid Ó Coileáin
CTVR, School of Computer Science and Statistics
Trinity College Dublin
Dublin 2, Ireland
Email: collindi@tcd.ie

Donal O'Mahony
CTVR, School of Computer Science and Statistics
Trinity College Dublin
Dublin 2, Ireland
Email: Donal.OMahony@cs.tcd.ie

***Abstract*—The Information Centric Networking (ICN) paradigm offers solutions to some of the functional and performance limitations of the current Internet architecture by offering secure, efficient and scalable mechanisms for the delivery of content to end-users. However, ICN architectures fail to provide adequate content accountability for the content distribution process. (We use accountability and feedback interchangeably to refer to the logging information collected as part of the content distribution process). In this paper we propose the Savant architecture, which is a scalable accountability framework for ICN content distribution. We outline our current work and some preliminary results, which demonstrate the viability and scalability of the Savant architecture. Finally, we briefly outline avenues for future work.**

I. Introduction

Cisco predict that the sum of all consumer video traffic (peer-to-peer (P2P), Video on Demand (VoD) and TV) across all networks (Internet, managed IP, mobile data) will surpass 1.6 zettabytes (2^{70} bytes) per annum by 2018, a threefold increase on 2013 traffic [1]. Much of this traffic is caused by the continual redistribution of popular time-shifted content (i.e., content viewed at a convenient time to the consumer e.g., VoD), which results in high costs for network operators and requires support from expensive CDN infrastructure [2]. As a result, content providers (i.e., entities that own or are licenced to sell or distribute content) are looking for more secure, efficient, scalable and accountable mechanisms for the delivery of content to end-users. The Information Centric Networking (ICN) paradigm offers solutions to some of these challenges by enabling content to *self-verify* i.e., a user can establish integrity, trust and provenance in content received from trusted or untrusted infrastructure.

However, no ICN architecture proposes a framework for providing content accountability despite depending on trusted and untrusted infrastructure to distribute content [3]. We define content accountability and feedback as the willingness of trusted or untrusted communicating entities to produce accurate and verifiable information about the content distribution process. This includes metrics about the performance of the content delivery infrastructure (e.g., CPU usage and RAM occupancy), end-user quality of experience (QoE) (e.g., buffering ratio and bitrate), end-user engagement (e.g., content views and advertising impressions) and user demographics

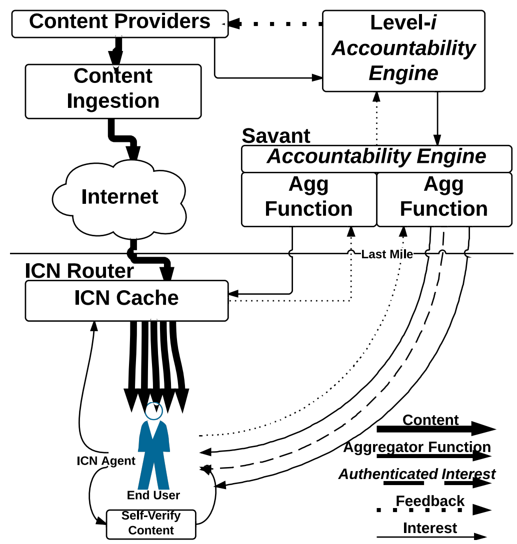

Fig. 1. The ICN architecture with the Savant framework component

(e.g., geographic location and device type). In fact, content accountability violates some ICN architecture principles related to privacy e.g., Named Data Networking (NDN) [4].

The Savant architecture proposes a scalable framework for monitoring and verifying accountability information collected in ICN networks by removing a content providers dependence on centralised infrastructure to consolidate large volumes of logging information. This is achieved using ICN principles and aggregator functions based on event processing languages [5] to limit the amount of accountability information collected and aggregated by Savant. In the remainder of this paper we outline the Savant architecture and associated benefits, some preliminary results from implementation and we discuss future work in the area.

II. Savant Overview

The principle components in the Savant architecture include the content provider, content ingestion process, client-side instrumentation, cache-side instrumentation, accountability engine and aggregator functions. Figure 1 provides an overview of the Savant framework. The accountability process is started when the content provider adds content for ingestion. The

QSHINE 2014, August 18-20, RHODES, Greece

DOI 10.4108/icst.qshine.2014.256405

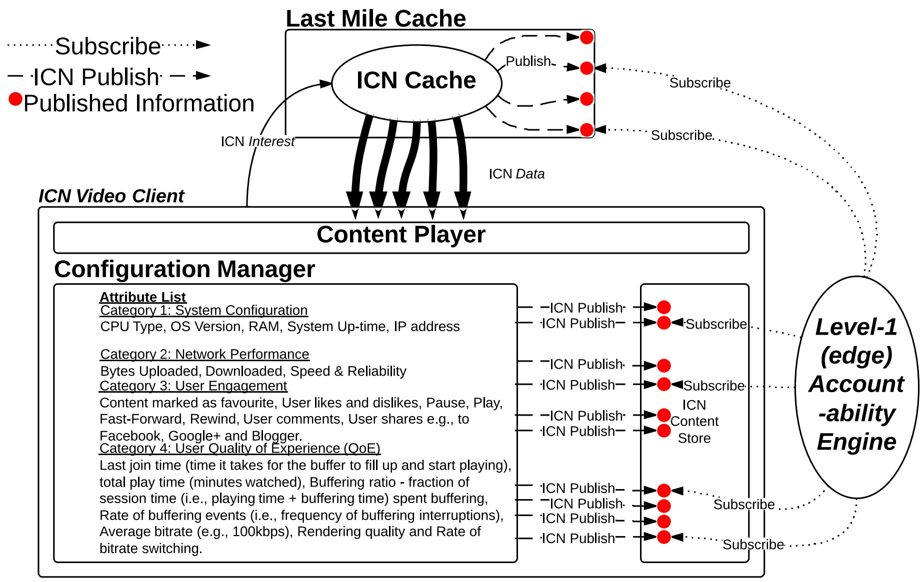

Fig. 2. The attribute values in categories identified here are published by ICN agents. The accountability engine subscribes to published data at ICN agents.

ingestion process prepares content for distribution to many different users, devices and networks, performing tasks such as transcoding, resolution conversion, content chunking and adding *metadata*. This metadata contains information about the content (e.g., genre, time-to-live, bitrate and content length) and the content publisher (e.g., name and public-key location). Moreover, metadata is fundamental in order to start the accountability process in client-side and cache-side instrumentation as it contains bootstrap information for the accountability engine. This metadata determines who will manage the collection and aggregation of published accountability information - either the content provider or a third party entity (such as a CDN). In this way, Savant offers content providers the ability to collect and aggregate their own accountability information, limiting or removing their dependence on third-party infrastructure for content distribution and feedback collection. Finally, additional aggregator functions (e.g., name and location) can also be added to the published content and metadata for customised feedback and collection at ICN caches or ICN clients, collectively referred to as *ICN agents*.

When published content passes through ICN agents, the local instrumentation library reads, interprets and publishes feedback information based on the metadata in the content received. The prefix part of the published data uses the hostID as the root namespace (used for routing). This published feedback information is collected by the accountability engine. The accountability engine is composed of geographically dispersed infrastructure located close to end-users (similar to CDN infrastructure), which has primary responsibility for collecting, aggregating, summarising and publishing feedback information collected from multiple ICN agents. This functionality is achieved using aggregator functions.

Aggregator functions (published as ICN content objects) form a key component of the accountability engine as they determine what feedback information to collect (depicted in Figure 2), the frequency to collect it and how to aggregate the data returned. This functionality is achieved using event processing languages [5], which run queries continuously as new data arrives. At their simplest level they filter, summarise and publish feedback data containing specified attributes. However, they are also responsible for more complex operations such as monitoring for patterns of complex events and notifying interested parties on occurrence. Additionally in circumstances where nodes are suspected of being malicious or causing Byzantine faults, aggregator functions can trigger the collection of additional accountability information from ICN agents. Furthermore, aggregator functions can also start ICN agents collecting hash chains of log entries in order to produce tamper-evident logs [2]. This is achieved using a similar mechanism to *Authenticated Interests* [6]. *Authenticated Interests* add commands and authentication tags (i.e., digital signatures or message authentication codes (MACs)) to ICN Interests to support applications efficiently running commands on remote fixtures [6]. This additional processing and complex verification is only required on ICN agents if aggregator functions on the accountability engine suspect faults with the content distribution process. Finally, Savant aggregates information hierarchically by adding additional levels (level-*i*) of accountability engine infrastructure, which can provide real-time scalable summary information to content providers for a large number of ICN users.

III. Benefits of Savant

There are several benefits offered by the Savant architecture, which constitute new functionality for content distribution systems. These benefits were identified as research challenges in information flow processing systems related to data uncertainty (imprecise or incorrect data), dynamic rule changing (dynamically changing aggregator functions), pull based (publish/subscribe) architectures as an IFP systems and support for compositional processing [5]. Savant's functionality is dependent on combining ICN publish/subscribe architectures with information flow processing systems [5]. This allows aggregator functions to request published accountability information from ICN agents, which can be increased or decreased dynamically. Moreover, this enables Savant to scale because it assumes that all clients and caches are good and initially collects accountability information based on that premise. Additionally, the overhead associated with collecting and analysing content feedback information is small, as observed in Section 4. Finally, Savant naturally supports gathering of non-repudiable evidence of ICN agent actions, which reduces the uncertainty associated with collected data.

IV. Preliminary Results

In a 2011 paper [7], a team at Conviva performed an analysis of data collected from over 1 million unique viewers, content providers, Internet service providers and CDNs in order to determine the impact video quality has on end-user engagement (i.e., the amount of time users interact with content). In order to prove the viability of the Savant architecture we partially replicated that experiment by gathering a subset of these video quality metrics for a small number of ICN users. We achieved this using CCNx, which is a software implementation of the Named Data Networking (NDN) ICN architecture. In order to gather these video quality metrics we made small modifications to the NDN Video [8] content distribution process to monitor join-time, session duration, buffering, rate of buffering events and average bitrate. These QoE metrics were collected from remote NDN clients by the Savant accountability engine, which was running on an Amazon Elastic Compute Cloud (EC2) micro-instance machine (i.e., Ubuntu-12.04.3-64 bit; memory: 613MB and disk: 8GB). We observed a linear increase in resource usage on the Savant Server for each additional NDN agent added. Based on these observations, we extrapolated that Savant could support 104 busy concurrent NDN agents for each Amazon EC2 micro-instance using about 61% of network resources used by one NDN Video content distribution session. Moreover Savant was able to identify problems with the NDN Video content distribution session. For example, Savant is the first system (to our knowledge) to consistently measure NDN Video join-time at over 10 seconds, which based on Conviva's analysis has a big impact on user engagement and user retention.

V. Future Work

The Savant framework reveals several avenues of research for supporting the content distribution process in ICN architectures. For example, a specific pattern of activity related to Byzantine faults or malicious caches distributing content that invariably leads to the accountability infrastructure redirecting the ICN client to another ICN cache resource. Furthermore, using accountability information collected it should be possible to support ICN-P2P content distribution because Savant knows the location of multiple copies of self-verifying content. Moreover, it should be possible for Savant to augment or replace ICN routing because it can act as the control plane for the content distribution process. This functionality is based on combining and interpreting the output of multiple complex event streams in information flow processing systems from clients, caches and accountability engine infrastructure. Finally, this functionality is based on standardising metadata and attribute information between content publishers, caching elements and consumers.

VI. Conclusion

In conclusion, Savant's current and envisaged future functionality should enable content providers to remove their dependence on a single entity (such as a CDN) for content distribution and feedback collection. This is facilitated by ICN publish/subscribe architectures and Savant's ability to consolidate and interpret logging information while maintaining feedback integrity.

VII. Acknowledgement

The authors would like to thank the anonymous reviewers for feedback that helped improve the paper. This research, a part of the Telecommunications Graduate Initiative, is funded by the Irish Higher Education Authority under the Programme for Research in Third-Level Institutes (PRTLI) and co-funded under the European Regional Development fund.

References

[1] Cisco. (2014) Cisco visual networking index: Forecast and methodology 2013–2018. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[2] P. Aditya, M. Zhao, Y. Lin, A. Haeberlen, P. Druschel, B. Maggs, and B. Wishon, "Reliable client accounting for p2p-infrastructure hybrids," in *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, ser. NSDI'12. Berkeley, CA, USA: USENIX Association, 2012, pp. 8–8. [Online]. Available: http://dl.acm.org/citation.cfm?id=2228298.2228309

[3] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. V. Katsaros, and G. C. Polyzos, "A survey of information-centric networking research," *Communications Surveys Tutorials, IEEE*, vol. 16, no. 2, pp. 1024–1049, Second 2014.

[4] L. Zhang, D. Estrin, J. Burke, V. Jacobson, J. Thornton, D. K. Smetters, B. Zhang, G. Tsudik, K. Claffy, D. Krioukov, D. Massey, C. Papadopoulos, T. Abdelzaher, L. Wang, P. Crowley, and E. Yeh, "Named Data Networking (NDN) Project NDN-0001," Tech. Rep., Oct. 2010. [Online]. Available: http://www.parc.com/content/attachments/named-data-networking.pdf

[5] G. Cugola and A. Margara, "Processing flows of information: From data stream to complex event processing," *ACM Comput. Surv.*, vol. 44, no. 3, pp. 15:1–15:62, Jun. 2012. [Online]. Available: http://doi.acm.org/10.1145/2187671.2187677

[6] J. Burke, P. Gasti, N. Nathan, and G. Tsudik, "Securing instrumented environments over content-centric networking: the case of lighting control," *CoRR*, vol. abs/1208.1336, 2012.

[7] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," in *Proceedings of the ACM SIGCOMM 2011 Conference*, ser. SIGCOMM '11. New York, NY, USA: ACM, 2011, pp. 362–373. [Online]. Available: http://doi.acm.org/10.1145/2018436.2018478

[8] D. Kulinski and J. Burke, "Ndn video: Live and prerecorded streaming over ndn," Technical Report, The NDN Project Team, Tech. Rep., 2012. [Online]. Available: http://http://named-data.net/publications/techreports/trstreaming/