AC^{3}N: Anonymous Communication in Content-Centric Networking

Gene Tsudik
Department of Computer Science
UC Irvine
Email: gts@ics.uci.edu

Ersin Uzun
Computer Science Laboratory
PARC
Email: ersin.uzun@parc.com

Christopher A. Wood
Department of Computer Science
UC Irvine
Email: woodc1@uci.edu

***Abstract*—Content-Centric Networking (CCN) is an emerging (inter-)networking architecture with the goal of becoming an alternative to the IP-based Internet. To be considered a viable candidate, CCN must at least have parity with existing solutions for confidential and anonymous communication, e.g., TLS, tcpcrypt, and Tor. ANDāNA (Anonymous Named Data Networking Application) was the first proposed solution that addressed the lack of anonymous communication in Named Data Networking (NDN)–a variant of CCN. However, its design and implementation led to performance issues that hinder practical use. In this paper we introduce AC^{3}N: Anonymous Communication for Content-Centric Networking. AC^{3}N is an evolution of the ANDāNA system that supports high-throughput and low-latency anonymous content retrieval. We discuss the design and initial performance results of this new system.**

I. Introduction

Network services and applications have undergone a tremendous transformation since their inception in the 1970s. Content distribution, instead of email and remote access to shared resources, has become the leading source of Internet traffic. For example, Netflix alone accounted for nearly 30% of all downstream Internet traffic in 2012 [4]. The number and popularity of such content-centric services are only expected to increase with the growth of data-intensive consumer applications and devices, leading to added pressure on network resources, increased congestion, and wasted bandwidth in today's Internet. Furthermore, with growing evidence of large-scale network packet interception and eavesdropping [9], consumer anonymity and privacy are becoming desired features of networking technologies and services.

Content Centric Networking (CCN) [14][16] is an emerging (inter-)networking architecture with the goal of becoming an alternative to the IP-based Internet. Two of its primary characteristics are that (1) content is named, addressable, and routable in the network, and (2) all content is signed by its producer. By the first property, a consumer who wishes to obtain content first issues a request (interest) for said content by name, which is then routed to the producer or a network entity (i.e., router) that is capable of satisfying the request. The corresponding content carrying the same name is then sent to the consumer along the reverse path. The second property enables content integrity and authenticity to be decoupled from where it is stored (cached) and how it is delivered to consumers. These two fundamental design characteristics permit content to be opportunistically cached throughout the network, thereby lowering congestion and improving overall bandwidth utilization.

The CCN architecture enables content access control via producer-specified forms of encryption or content name obfuscation. However, support for consumer and producer anonymity is not a standard feature. ANDāNA (Anonymous Named-Data Networking Application) [11] is the initial attempt to provide anonymous communication in Named Data Networking (NDN)[2], an early variant of CCN. Inspired by Tor [12][3], it uses onion-like concentric encryption to wrap interests for content that are gradually decrypted and forwarded by participating anonymizing routers. Along the return path, content is wrapped in layers of encryption as it flows from the satisfying entity to the consumer.

In this work we present an improvement to the ANDāNA design. Our approach, henceforth referred to as AC^{3}N (Anonymous Communication for Content-Centric Networking), addresses some of the performance and anonymity problems with ANDāNA. The design of AC^{3}N relies only on the underlying network's ability to pull uniquely named content by name (via interests) and does not depend on any other features, e.g., in-network caching. This makes AC^{3}N applicable in any CCN incarnation, e.g., CCNx [1] and NDN [2].

The remainder of this paper focuses on the shortcomings of ANDāNA and the improved AC^{3}N design which addresses these pitfalls. We also report on its performance as an application built on top of CCNx 1.0 citeccnx. Our initial experiments and analysis indicate that AC^{3}N can support high-throughput, low-latency, unidirectional and bidirectional traffic over CCNs with anonymity guarantees equivalent to Tor.

II. Preliminaries

This section overviews the properties of the CCN architectures that are relevant to anonymous communication. We then assess the ANDāNA design and identify certain engineering shortcomings as well as anonymity flaws that are remedied by AC^{3}N.

A. CCN Overview

Content distribution in CCN follows a *pull model* whereby content is requested by consumers by name instead of by location (e.g., an IP address). These requests, called *interests*, contain the name of the desired content rather than its location. Although an *interest* is intended to carry a meaningful (human-readable) URI-like name, it can in fact carry an arbitrary string corresponding, such as encoded binary data.

An interest is routed based on the specified name over a sequence of routers, each of which keeps state of the forwarded interest. An interest might get routed to a producer who would reply with the requested content. Alternatively, the requested content might be found in a cache of an intermediate router along the consumer-to-producer path. The latter can occur because each router is expected, though not mandated, to opportunistically cache every content object it forwards to consumers. The bidirectional flow of interests and content in the network guarantees that all traffic flow is symmetric; A single content object is always returned in response to an interest along the same consumer-to-producer path.

B. ANDāNA *Highlights*

To motivate AC^{3}N, we first re-examine ANDāNA, the first anonymous communication tool designed for CCNs. In ANDāNA, just as in Tor, interests and content objects traverse *circuits* (paths) of anonymizing routers (ARs). The ARs in a circuit are chosen by the consumer. Before interests are issued by a consumer, their name is first *wrapped* in concentric layers of encryption. Each "layer" contains a routable name prefix for the next hop (AR) in the circuit and the underlying encrypted layers. Each AR decrypts their layer of the name to obtain the next routable prefix in the circuit and corresponding layer (i.e., it "unwraps" its layer of encryption), and then forwards the interest with the new name accordingly. Upon the receipt of content objects in the reverse path, each AR will encrypt the entire content object and forward the "wrapped" result to the next downstream hop. The consumer then recovers the content object by iteratively decrypting each layer of encryption surrounding the content object.

Unlike Tor, ANDāNA does not support persistent anonymous circuits between consumers and producers. Rather, ephemeral (one-time) circuits are created as the interest is sequentially decrypted (using public-key decryption) and forwarded. Similarly, content is encrypted with a public-key on the reverse path. Circuit state information in each AR is only maintained in the symmetric (session-based) variant ANDāNA.

In the symmetric variant of ANDāNA, state information, consisting of a unique session identifier and symmetric key used for interest and content decryption and encryption, is established in each anonymizing router using a standard three-way handshake protocol. The use of symmetric encryption removes the computational burden of public key encryption. However, the ANDāNA design requires that the session identifier be sent in the clear for every interest, which allows an adversary to link interests and content packets, thus enabling deanonymization attacks against consumers Furthermore, the handshake procedure wastes consumer bandwidth and time, especially in the case of short-term communication.

C. Identified Issues

The primary motivation for AC^{3}N is to attain the same anonymity guarantees as the public-key variant of ANDāNA with better versatility and performance. Although ANDāNA includes a symmetric (session-based) variant as a more efficient alternative, it does not provide interest and content object unlinkability. The ability to link interest and content objects can subsequently lead to consumer and producer linkability, which can immediately violate anonymity.

For example, suppose that an adversary $\mathcal{A}$ eavesdrops on incoming and outgoing interests for a particular AR. By looking at the traffic patterns, $\mathcal{A}$ can link incoming and outgoing session IDs. A variant of this adversary was studied in the context of Tor by Murdoch and Danezis in [17] and was shown to be quite successful. We believe that the same attack could be augmented to apply to ANDāNA. In particular, repeating this attack at each AR in a circuit can result in deanonymization of both the consumer and producer.

The use of application and environment contextual information has been investigated in [13], where side-channel and environment information (e.g., deterministic behavior of an AR always forwarding a packet after unwrapping an interest received from a downstream neighbor) is used to quantify the *degree of unlinkability*. Furthermore, regardless of how linkability information is acquired, it has been shown that it can degrade consumer and producer anonymity beyond that attainable by general traffic analysis [18].

Since most relevant literature focuses on mix-based anonymizing services akin to Tor, upon which ANDāNA was designed, it is clear that all linkability problems studied in the context of Tor are also applicable to symmetric variant of ANDāNA. This is why one of the key goals of AC^{3}N is to attain the same anonymity guarantees as the public key variant of ANDāNA, which has no linkability issues, while still providing more efficient support for low-latency, high-throughput, and bidirectional traffic as compared to the symmetric variant of ANDāNA.

III. AC^{3}N Design

A. Circuit and Session Establishment

Similar to Tor [12] and ANDāNA, anonymizing routers (ARs) and circuits are at the core of AC^{3}N. As previously mentioned, a *circuit* is a sequence of ARs through which upstream interests and downstream content objects flow. Circuits are established for long-term sessions, i.e., they are not ephemeral. ARs in a circuit serve two purposes: (1) decapsulate (decrypt) and forward encrypted interests, and (2) encapsulate (encrypt) content objects using previously acquired or agreed upon keys and forward them downstream.

Consumers generate interests wrapped in several layers of encryption and receive content objects also wrapped in several layers of encryption that it can decrypt. Each AR is an

application running on router, and therefore technically serves as the producer for each downstream AR in the AC^{3}N circuit. The standard CCN communication model suggests that such content *must be signed* with a trusted private key. However, AC^{3}Nstrays from this requirement and uses symmetric-key MACs for more efficient authenticity checks.

To increase interest and content throughput, circuit sessions are established and initialized with forward-secure symmetric keys used for content encryption and MAC tag generation and verification. The complete set of session state information, which is established for n ARs $r_1, \ldots, r_n$ in a circuit, is as follows:

- Session IDs $\mathsf{Session}_i^0$ and session initialization vectors (IVs) $\mathsf{SessionIV}_i^0$,
- Interest interest encryption keys and pseudorandom generator seed values $\mathsf{EncryptionIV}_i$, and
- Pairwise MAC keys M_{k_i} between adjacent ARs and the consumer (used to tag and verify content).

The $\mathsf{SessionIV}_i^0$ element is a counter of capacity 2^κ. Its use is described below. This state is established by running a three-way handshake protocol between the consumer and each AR. Both parties in this protocol contribute appropriate randomness for each element of the state for forward secrecy. This ensures that session information cannot be compromised if one of the AR private keys is also compromised.

After a circuit and its session information have been established, all subsequent traffic is protected via a CCA-secure symmetric scheme [15]. The encryption and MAC keys for AR r_j are indexed via $\mathsf{SessionIndex}_i^j$, the dynamic session index (identifier) sent in the cleartext along with the encrypted interest. To provide unlinkability, the session index is "advanced" from $\mathsf{SessionIndex}_i^j$ to $\mathsf{SessionIndex}_i^{j+1}$, after each new interest is received and forwarded, using a one-way and collision-resistant hash function $H(\cdot)$. The transfer functions for this variable are:

$$\mathsf{SessionIndex}_i^{j+1} = H(\mathsf{SessionIV}_i^j + \mathsf{SessionIndex}_i^j) \quad (1)$$

$$\mathsf{SessionIV}_i^{j+1} = 1 + \mathsf{SessionIV}_i^j \mod (2^\kappa). \quad (2)$$

Note that $\mathsf{SessionIV}_i^j$ is kept private. This means that an eavesdropper cannot guess the session index based on any past observations of interest and content exchanges between a consumer and AR. This dynamic session index is the key element that enables AC^{3}N to provide unlinkability.

State initialization in AC^{3}N is separate in time from circuit usage, i.e., it uses a handshake routine to initialize state. However, our design does not preclude on-line state establishment. For example, the first wrapped interest issued by a consumer for a new circuit could be overloaded to include all of the state establishment information in addition to the associated interest information. Note that this would not achieve the same forward secrecy guarantees as the handshake-based approach. Moreover, to be useful, these initial interests must not be noticeably different from other interests for the same circuit. As a result, all subsequent interests would have to be appropriately padded. This approach is therefore not used due to the potentially severe communication cost and added complexity.

B. AC^{3}N *Circuit Usage*

AC^{3}N circuits are used the same way as in ANDāNA. Specifically, a consumer c wraps an interest for a sequence of ARs and forwards it towards the first AR. This wrapping procedure is identical to the symmetric variant of ANDāNA wherein symmetric interest encryption keys are used to concentrically encrypt interest names. The appropriate session index is prepended to each encrypted name and then subsequently advanced as per equation (1). Also, each encrypted interest also includes a timestamp to mitigate replay attacks.

Content encryption at an AR uses XOR-based symmetric encryption with a key stream generated by a cryptographically secure pseudorandom generator with input $\mathsf{EncryptionIV}_i$. This random seed is advanced as a counter (similarly to the $\mathsf{SessionIV}_i$) so that the key stream is a fresh pseudorandom bit string for each content object.

The commutative property of XOR allows the consumer to decrypt each layer of the content in any arbitrary order. One additional benefit of this form of encryption is that it permits the key stream to be precomputed offline. It does, however, introduce the probability of improperly computed key streams, which will result in corrupt ciphertext. Also, after encrypting a content object, each AR will compute a MAC tag σ before forwarding the result to the next downstream hop. ARs and the consumer share pairwise MAC keys so that each can verify the tag of a content object upon receipt.

IV. Performance Assessment

ANDāNA was originally implemented in C using the CCNx 0.8x library. As outlined in the section II, the CCNx protocol and implementation has changed significantly since this original work. Thus, to bring the evaluation up to speed with existing technology, both ANDāNA and AC^{3}N were implemented over the CCNx 1.0. All experiments were conducted on VMs running Ubuntu 14.04 LTS. Each host was equipped with an Intel(R) Core(TM) i5-3427U CPU at 1.80GHz with 8GB of main memory. Also, we use the public key variant of ANDāNA, since it provides the same functionality as AC^{3}N.

In this work, we consider the most important metrics for performance to be (a) interest-content latency and (b) and throughput. To assess these metrics, we conducted the following simple experiment. Let $\mathsf{Circ} = R_1, \ldots, R_n$ be a circuit of length $n-1$ with n AR nodes. A client C is connected to R_1, and C wishes to retrieve content from producer P connected to R_n. Thus, the complete path is $C, R_1, \ldots, R_n, P$. To obtain content, C issues a random interest that can be satisfied by P through Circ. Such an interest is issued once every t seconds and the RTT is recorded. To compute the average latency, the average of all observed RTTs is computed. To compute the maximum throughput, the total number of content bytes received is divided by the total time to send a large amount of back-to-back interests without delay. The RTT and throughput

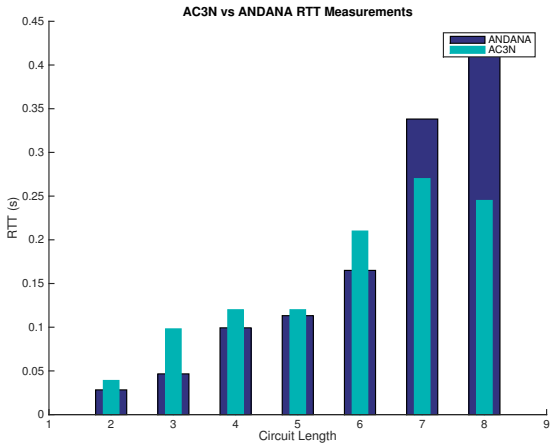

Fig. 1. Unidirectional RTT measurements.

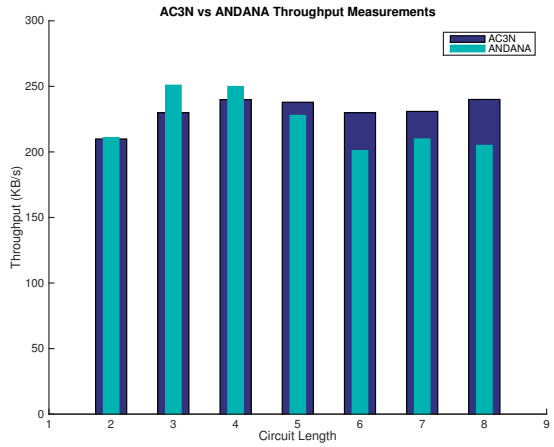

Fig. 2. Unidirectional throughput measurements.

measurements as a function of the circuit length are shown in Figures 1 and 2.

V. Related Work

Beyond ANDāNA there has not been much prior work on anonymous communication in content-centric networks. One approach to consumer privacy is proposed in [6]. Rather than using encryption, it requires content producers to mix sensitive information with so-called "cover" content. In addition to forcing producers to participate in consumer anonymity, all cover content must also be stored for a finite length of time. Furthermore, it does not provide protection against malicious producers and does not offer consumer-producer unlinkability. The use of covert channels for communication was first explored by Ambrosin et al. [5]. Compagno et al. [10] discuss deanonymization attacks on consumers using geo-location information. Tourani et al. [19] present a system which obviates trivial censorship techniques based on interest and content names in NDN. Their design is not based on Tor. Chen et al. [8] developed a high-speed onion routing system called HORNET. Their design does not require any per-flow state in each AR and only uses symmetric encryption, which allows it to process anonymous traffic at 93 GB/s. However, HORTEN is targeted for future internet architectures like SCION [7] that are not focused on content delivery.

VI. Conclusions and Future Work

This paper presented AC^{3}N–an application-layer anonymizing service for CCN. Its simple design relies only upon the fundamental interest-content request paradigm of information retrieval. Experiments indicate that AC^{3}N can support high-throughput, low-latency, unidirectional and bidirectional traffic over CCNs with anonymity guarantees equivalent to Tor.

References

[1] CCNx. http://ccnx.org/.
[2] Named-data networking. http://named-data.net/.
[3] Tor project website. https://www.torproject.org/.
[4] V. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z.-L. Zhang. Unreeling netflix: Understanding and improving multi-cdn movie delivery. In *INFOCOM, 2012 Proceedings IEEE*, pages 1620–1628, March 2012.
[5] M. Ambrosin, M. Conti, P. Gasti, and G. Tsudik. Covert ephemeral communication in named data networking. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*, pages 15–26. ACM, 2014.
[6] S. Arianfar, T. Koponen, B. Raghavan, and S. Shenker. On preserving privacy in content-oriented networks. In *Proceedings of the ACM SIGCOMM workshop on Information-centric networking*, pages 19–24. ACM, 2011.
[7] D. Barrera, R. M. Reischuk, P. Szalachowski, and A. Perrig. The scion internet architecture.
[8] C. Chen, D. E. Asoni, D. Barrera, G. Danezis, and A. Perrig. Hornet: High-speed onion routing at the network layer. *arXiv preprint arXiv:1507.05724*, 2015.
[9] CNN. Latest nsa leaks point finger at high-tech eavesdropping hub in uk, 2013. http://www.cnn.com/2013/12/20/world/europe/nsa-leaks-uk/.
[10] A. Compagno, M. Conti, P. Gasti, L. V. Mancini, and G. Tsudik. Violating consumer anonymity: Geo-locating nodes in named data networking.
[11] S. DiBenedetto, P. Gasti, G. Tsudik, and E. Uzun. Andana: Anonymous named data networking application. In *Network and Distributed System Security - NDSS*. The Internet Society, 2012.
[12] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, pages 21–21, Berkeley, CA, USA, 2004. USENIX Association.
[13] M. Franz, B. Meyer, and A. Pashalidis. Attacking unlinkability: The importance of context. In N. Borisov and P. Golle, editors, *Privacy Enhancing Technologies*, volume 4776 of *Lecture Notes in Computer Science*, pages 1–16. Springer Berlin Heidelberg, 2007.
[14] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking named content. In *Proc. ACM CoNEXT 2009*, pages 1–12, Dec. 2009.
[15] J. Katz and Y. Lindell. *Introduction to modern cryptography: principles and protocols*. CRC Press, 2007.
[16] M. Mosko. Ccnx 1.0 protocol specification roadmap. 2013.
[17] S. Murdoch and G. Danezis. Low-cost traffic analysis of tor. In *Security and Privacy, 2005 IEEE Symposium on*, pages 183–195, May 2005.
[18] S. Schiffner and S. Clauß. Using linkability information to attack mix-based anonymity services. In *Proceedings of the 9th International Symposium on Privacy Enhancing Technologies*, PETS '09, pages 94–107, Berlin, Heidelberg, 2009. Springer-Verlag.
[19] R. Tourani, S. Misra, J. Kliewer, S. Ortegel, and T. Mick. Catch me if you can: A practical framework to evade censorship in information-centric networks. In *Proceedings of the 2nd International Conference on Information-Centric Networking*, pages 167–176. ACM, 2015.