Security and Privacy Challenges in Content-Centric Networks

Christopher A. Wood

Ph.D. Defense
Donald Bren School of Information and Computer Sciences
University of California, Irvine
November 6, 2016
Outline

1. Overview

   - Access Control Dimensions
   - Interest-Based Access Control
   - Revocation and Network Eviction

   - Privacy Parity
   - Frequency Analysis Attacks

4. Conclusion

Christopher A. Wood
Security and Privacy Challenges in Content-Centric Networks
Content Retrieval

Simple enough, yet:

How does Jack discover example.com?

How does Jack actually ask for /foo/bar?

How do we keep /foo/bar secure (and private) in transit?
Simple enough, yet:

- How does Jack discover example.com?
- How does Jack actually ask for /foo/bar?
- How do we keep /foo/bar secure (and private) in transit?
Content Retrieval: Expanded

Jack

TCP SYN
TCP SYN+ACK
TCP ACK
TLS Hello
TLS Hello, KeyExchange
TLS KeyExchange, Finished
TLS Finished
HTTP GET /foo/bar
example.com
2001:…
DNS
QUERY
example.com
A: A.B.C.D
AAAA: 2001:…

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
Content Retrieval: To Each Their Own

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
Shortcomings

Main problems:

- Wasted bandwidth
- Competing congestion control
- Excessive client-server round trips
- Heavy reliance on complicated (and fragile) public key infrastructure
Shortcomings

Main problems:
- Wasted bandwidth
- Competing congestion control
- Excessive client-server round trips
- Heavy reliance on complicated (and fragile) public key infrastructure

Today’s Internet:
- Designed for end-to-end communication
- Not designed with security or privacy in mind
- Encourages channel or transport encryption
- Not suited to content distribution
Content-Centric Networking

Data is named
Content-Centric Networking

1. Data is named
2. Name-to-data binding is cryptographic
Content-Centric Networking

1. Data is named
2. Name-to-data binding is cryptographic
3. Consumers request content via interests carrying data names
1. Data is named
2. Name-to-data binding is cryptographic
3. Consumers request content via interests carrying data names
4. Routers forward interests towards producers based on their names
Content-Centric Networking

1. Data is named
2. Name-to-data binding is cryptographic
3. Consumers request content via interests carrying data names
4. Routers forward interests towards producers based on their names
5. Routers forward data in content objects towards consumers using router state
   - Content may be cached if desired
Content-Centric Networking

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
Content-Centric Networking

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
Content-Centric Networking
Content-Centric Networking

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks

12
IP Content Retrieval: To Each Their Own

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
CCN Interest Forwarding

\[ I(N) = /edu/ics/uci/woodc1/files/phd-defense.pdf \]
CCN Interest Forwarding

\[ I(N) = \text{/edu/ics/uci/woodc1/files/phd-defense.pdf} \]
CCN Interest Forwarding

CCN Forwarder

I(N) = /edu/ics/uci/woodc1/files/phd-defense.pdf

routable prefix unroutable suffix
CCN Interest Forwarding

I(N) = /edu/ics/uci/woodc1/files/phd-defense.pdf

- routable prefix
- unroutable suffix

Christopher A. Wood
Security and Privacy Challenges in Content-Centric Networks
CCN Content Forwarding

CCN Forwarder

C(N) → Forward → CS

Verify
Content

PIT Hit → Insert

PIT

Lookup

PIT Miss → Drop

X

C(N) → FIB

Verify
Content

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks

21
CCN Content Forwarding

CCN Forwarder

C(N) → Forward

CS

PIT Hit
Insert

PIT

Lookup

FIB

PIT Miss
Drop

C(N) → C(N)

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
CCN Challenges

Access control: protecting data from unauthorized consumers

CCN Challenges

**Access control:** protecting data from unauthorized consumers


**Privacy:** preventing data-to-consumer linkability

CCN Challenges

Access control: protecting data from unauthorized consumers

Privacy: preventing data-to-consumer linkability

Availability: mitigating or deterring denial of service attacks
Outline

1. Overview

   - Access Control Dimensions
   - Interest-Based Access Control
   - Revocation and Network Eviction

   - Privacy Parity
   - Frequency Analysis Attacks

4. Conclusion
Question #1: How can we ensure that only authorized users are able to read content?

1. Encrypt content and give decryption keys to authorized users
2. Disable caching, require interest authentication
3. Make interest names only derivable by authorized users

Caches are still okay!
Access Control Problem

**Question #1:** How can we ensure that only *authorized users* are able to read content?

1. Encrypt content and give decryption keys to authorized users
2. Disable caching, require interest authentication
3. Make interest names only derivable by authorized users
   - Caches are still okay!
Question #1: How can we ensure that only authorized users are able to read content?

1. Encrypt content and give decryption keys to authorized users
2. Disable caching, require interest authentication
3. Make interest names only derivable by authorized users
   - Caches are still okay!

Question #2: How do we take access away (revocation)?
Content-Based Access Control (CBAC)

Encrypt the content and give only the keys to authorized consumers
Content-Based Access Control (CBAC)

Encrypt the content and give only the keys to authorized consumers

Hot topic in literature:

- ...
Interest-Based Access Control (IBAC)

Only authorized consumers can request content
Interest-Based Access Control (IBAC)

Only authorized consumers can request content

- \( N' = f(k, N) \) for some function \( f \) and (group \( G \)) secret key \( k \)
- \( A \) without (\( G \)) secret cannot generate obfuscated names
Interest-Based Access Control (IBAC)

Only authorized consumers can request content

- \( N' = f(k, N) \) for some function \( f \) and (group \( G \)) secret key \( k \)
- \( A \) without (\( G \)) secret cannot generate obfuscated names

Related work:

- ...

Christopher A. Wood
Obfuscation Variants

At least two obfuscation variants:

- Encryption: \( N' = \text{Enc}(k, N) \)
- Keyed Hash: \( N' = H(k, N) \)
Obfuscation Variants

At least two obfuscation variants:

- Encryption: \( N' = Enc(k, N) \)
- Keyed Hash: \( N' = H(k, N) \)

The routable prefix is unaffected
**Question**: What if we want group-based access control?

*Answer*: Consumers in group $G_i$ share encryption key $k_{G_i}$.

*Question*: How does a producer identify $k_{G_i}$?

*Answer*: Include identifier $ID_{G_i} = H(k_{G_i})$ in interests.

*Question*: How can we prevent linkability of different interests with $ID_{G_i}$?

*Answer*: Encrypt identifiers using producer's public key $pk_P$: $\bar{ID}_{G_i} = Enc(pk_P, ID_{G_i})$. 
Supporting Multiple Groups

**Question:** What if we want group-based access control?

**(One) Answer:** Consumers in group $G_i$ share encryption key $k_{G_i}$

---

Christopher A. Wood

*Security and Privacy Challenges in Content-Centric Networks* 29
Supporting Multiple Groups

**Question**: What if we want group-based access control?

**(One) Answer**: Consumers in group $G_i$ share encryption key $k_{G_i}$

**Question**: How does a producer identify $k_{G_i}$?

**Answer**: Include identifier $ID_{G_i} = H(k_{G_i})$ in interests
Supporting Multiple Groups

**Question**: What if we want group-based access control?

(One) **Answer**: Consumers in group \( \mathbb{G}_i \) share encryption key \( k_{\mathbb{G}_i} \).

**Question**: How does a producer identify \( k_{\mathbb{G}_i} \)?

(One) **Answer**: Include identifier \( \text{ID}_{\mathbb{G}_i} = H(k_{\mathbb{G}_i}) \) in interests.
Supporting Multiple Groups

**Question**: What if we want group-based access control?

*(One) Answer*: Consumers in group $G_i$ share encryption key $k_{G_i}$

**Question**: How does a producer identify $k_{G_i}$?

*(One) Answer*: Include identifier $ID_{G_i} = H(k_{G_i})$ in interests

**Question**: How can we prevent linkability of different interests with $ID_{G_i}$?
Supporting Multiple Groups

**Question:** What if we want group-based access control?

**(One) Answer:** Consumers in group $G_i$ share encryption key $k_{G_i}$

**Question:** How does a producer identify $k_{G_i}$?

**(One) Answer:** Include identifier $ID_{G_i} = H(k_{G_i})$ in interests

**Question:** How can we prevent linkability of different interests with $ID_{G_i}$?

**(One) Answer:** Encrypt identifiers using producer’s public key $pk^P$:

$$\bar{ID}_{G_i} = Enc(pk^P, ID_{G_i})$$
On-Path $\mathcal{A}$

On-path $\mathcal{A}$ can:

- Observe obfuscated interests in transit
- Replay interests *for cached content*
On-Path $\mathcal{A}$

On-path $\mathcal{A}$ can:

- Observe obfuscated interests in transit
- Replay interests *for cached content*

Replay prevention requires:

- Nonces and timestamps
- Consumer authentication information

\[ \text{payload} = (\text{ID}_G, r, t, \sigma = \text{Sign}_{sk}(N') || \text{ID}_G || r || t) \]
On-Path $\mathcal{A}$

On-path $\mathcal{A}$ can:

- Observe obfuscated interests in transit
- Replay interests *for cached content*

Replay prevention requires:

- Nonces and timestamps
- Consumer authentication information

$$\text{payload} = (\text{ID}_G, r, t, \sigma = \text{Sign}_{sk}^{s}(N' || \text{ID}_G || r || t))$$
Question: How are interests for cached content authorized?
Interest Authentication

**Question:** How are interests for cached content authorized?

**Answer:** Verify interest authenticators
Interest Authentication

**Question:** How are interests for cached content authorized?

**Answer:** Verify interest authenticators

**Question:** What verification key?
Interest Authentication

**Question**: How are interests for cached content authorized?

**Answer**: Verify interest authenticators

**Question**: What verification key?

**Answer**: Producer-supplied, per-content key(s)
Question: How are interests for cached content authorized?

Answer: Verify interest authenticators

Question: What verification key?

Answer: Producer-supplied, per-content key(s)

**Authorized Content-Key Binding (ACKB):** Cached content protected under IBAC must reflect the verification key associated with the authorization policy.
ACKB in Action (Part 2)

\[ C r_2 \in G_i \]

\[ ID_{G_i} \leftarrow H(k_{G_i}) \]
\[ N' \leftarrow /prefix/Enc(k_{G_i}, \text{Suffix}(N, \text{prefix})) \]
\[ r_2 \overset{s}{\leftarrow} \{0, 1\}^\kappa, t_2 \leftarrow \text{CurrentTime()} \]
\[ \sigma \leftarrow \text{Sign}_{sk_{G_i}}(N' || ID_{G_i} || r_2 || t_2) \]
\[ \text{Payload} := (ID_{G_i}, r_2, t_2, \sigma) \]

\[ I[N']_2 := (N', \text{payload}) \]

\[ \text{Verify}_{pk_{G_i}}(\sigma), r_2 \text{ and } t_2 \]

\[ CO[N'] := (N', \text{data}, pk_{G_i}) \]
**System Setup:** 2.8 GHz Intel Core i7 CPU and 16GB of 1600 MHz DDR3 RAM running Ubuntu 14.04

**Signature details:** ElGamal signature algorithm, 1024-bit keys
IBAC Review

- Helps name privacy
- Replay attack prevention requires consumer authentication by trusted routers
- Best coupled with CBAC to obviate replay attacks
Question #1: How can we ensure that only *authorized users* are able to access (the body of) a content object?

Question #2: How do we take access away (revocation)?
Revocation

Revocation requires two steps:

- Rotating group encryption keys (easy)
- Erasing stale content from network caches (hard)

Related work:
Revocation

Revocation requires two steps:

- Rotating group encryption keys (easy)
- Erasing stale content from network caches (hard)

Related work:

Erasure Requirements

1. Erase messages must be authenticated (easy)
2. Erase messages must be distributed to all routers which may have potentially cached the content (hard)
Authenticating Erase Messages

Several things to consider:

1. Authentication should not cause a DoS on routers
2. Deletion is an idempotent operation
Authenticating Erase Messages

Several things to consider:

1. Authentication should not cause a DoS on routers
2. Deletion is an idempotent operation

Solution: use pre-image resistant hash functions

1. Generate $x \leftarrow \{0, 1\}^\lambda$
2. Compute $y = H(x)$
3. Distribute $y$ with publish content
4. Publish $x$ with erase message
Several things to consider:

1. Producers have no knowledge about where content may be stored
2. Not all routers may cache content
3. Cannot address routers directly
Distributing Erase Messages

Several things to consider:

1. Producers have no knowledge about where content may be stored
2. Not all routers may cache content
3. Cannot address routers directly

Reminiscent of IP traceback:

- ...
Hierarchy of distribution mechanisms:

1. Mark Interests
2. **Maintain lossless forwarding histories** (content fingerprint cache)
3. **Maintain lossy forwarding histories** (fingerprint bloom filters)
4. Broadcast (reverse flood) erase messages
Hierarchy of distribution mechanisms:

1. Mark Interests
2. **Maintain lossless forwarding histories** (content fingerprint cache)
3. **Maintain lossy forwarding histories** (fingerprint bloom filters)
4. Broadcast (reverse flood) erase messages

Diagram:

- E[a/b, d]
- E[a/b, d]
- C[a/b], {F0, F2}
- F0
- F1
- F2
- H0
- H1
- H2
- F3
- F4
- F5
- F6
- F7
- CS
- (1) check cache
- (2) check forwarding histories
- (3) forward erase message for matches
- E[a/b, d]
History-Based Forwarding

\[ E(N, d) \]

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
History-Based Forwarding

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
History-Based Forwarding

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks 44
Forwarder History Saturation

Consumer-facing router (BF of size 4GB, 3,200 Cps):
Forwarder History Saturation

Consumer-facing router (BF of size 4GB, 3,200 Cps):

- Lossless history: saturation in 12 hours
Forwarder History Saturation

Consumer-facing router (BF of size 4GB, 3,200 Cps):
- Lossless history: saturation in 12 hours
- Lossy history: saturation in 24 hours (false probability ceiling of $10^{-32}$)
Experimental Results

Goal: determine router and network overhead
- Consumers requesting at 10 interest/second
- Producers deleting a random 50% of their content
Experimental Results

Goal: determine router and network overhead
- Consumers requesting at 10 interest/second
- Producers deleting a random 50% of their content

Topology: large DFN and ATT topologies
Outline

1 Overview

2 Access Control [CCNC 2014, IFIP Networking 2015, ICN 2015, IFIP Networking 2016]
   - Access Control Dimensions
   - Interest-Based Access Control
   - Revocation and Network Eviction

3 Privacy [CCNC 2016, WPES 2016, IFIP Networking 2017 (x2), ICN 2017, LCN 2017]
   - Privacy Parity
   - Frequency Analysis Attacks

4 Conclusion

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
IP Privacy

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
IP Privacy

What is leaked?

- Source and destination addresses and ports
- Packet sizes and timing
CCN Privacy
CCN Privacy

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
CCN Privacy

What is leaked?

- Relative packet locations
- Packet timing and sizes
- Producer identities
- Interest names (and equality)

Related work:

- ...
Privacy Parity

CCN privacy < IP privacy
CCN privacy < IP privacy ?

- Interests for same content can be correlated
- Interest names reveal information about content
- Content carries explicit names
- Location of originator not (always) apparent

To what extent is (encrypted) data private?
CCN privacy < IP privacy?

What is the “delta”?

- Interests for same content can be correlated
- Interest names reveal information about content
- Content carries explicit names
- Location of originator not (always) apparent
Privacy Parity

CCN privacy < IP privacy?

What is the “delta”? 
- Interests for same content can be correlated
- Interest names reveal information about content
- Content carries explicit names
+ Location of originator not (always) apparent

To what extent is (encrypted) data private?
Data Privacy Threat Model

Several attacks:

- Correlation: learn when two requests correspond to same content
- Identification: learn when specific content was requested
- Leakage: learn anything from a request or response
Data Privacy Threat Model

Several attacks:
- Correlation: learn when two requests correspond to same content
- Identification: learn when specific content was requested
- Leakage: learn anything from a request or response

Adversaries:
- Eavesdropper: a passive interceptor
- On-path HbC: router that forwards interest and content packets
- Distributed: at least two on-path – one near producer, one near consumer
- Active & Scary: as above, also generates its own probes
Weak and Strong Privacy

**Weak**: $A$ can not learn anything from a request or response, but can correlate packets

- Protect responses with IND-KPA secure encryption
- Protect requests with deterministic cryptographic PRF (not length preserving)
Weak and Strong Privacy

**Weak**: $A$ can not learn anything from a request or response, but can correlate packets

- Protect responses with IND-KPA secure encryption
- Protect requests with deterministic cryptographic PRF (not length preserving)

**Strong**: $A$ can not learn, identify, or correlate

- Protect requests and responses with IND-CCA encryption
Weak Privacy and Deterministic Encryption

Deterministic encryption is subject to **frequency analysis attacks**

- $\mathcal{A}$ has a priori information about plaintext content popularity
- $\mathcal{A}$ observes frequency of requests for encrypted content
- $\mathcal{A}$ aligns frequency distribution with popularity distribution

Related work:


...
Deterministic encryption is subject to **frequency analysis attacks**

- $\mathcal{A}$ has a priori information about plaintext content popularity
- $\mathcal{A}$ observes frequency of requests for encrypted content
- $\mathcal{A}$ aligns frequency distribution with popularity distribution

**Related work:**

- ...
Active & Scary Adversary

Nefarious ISPs, nation states, etc:
- To what extent does auxiliary information accuracy matter?
- To what extent does universe size matter?
Active & Scary Adversary

Nefarious ISPs, nation states, etc:
- To what extent does auxiliary information accuracy matter?
- To what extent does universe size matter?

No Auxiliary Information

Perfect Auxiliary Information
Auxiliary Information Gap
Distributed Adversary

Access point, enterprise network middlebox, compromised transit router, etc.

- Where does the adversary have the best chance at succeeding?
- To what extent does caching dampen request frequency?
Access point, enterprise network middlebox, compromised transit router, etc.

- Where does the adversary have the best chance at succeeding?
- To what extent does caching dampen request frequency?
Distributed Adversary

**Edge vs Inner Router**

- **Graph**
  - X-axis: Content Rank
  - Y-axis: Percentage Correct [%]
  - Two lines:
    - Red: Edge routers
    - Blue: All routers

**Cache Presence**

- **Graph**
  - X-axis: Content Rank
  - Y-axis: Match Ratio
  - Three lines:
    - Red: \( p_c = 1.0 \)
    - Blue: \( p_c = 0.5 \)
    - Green: \( p_c = 0.0 \)
Takeaways

- Large content universe may render attack ineffective
  - Dynamic content is not susceptible to frequency analysis
- Caching helps and hurts privacy - use it sparingly
Outline

1 Overview

2 Access Control [CCNC 2014, IFIP Networking 2015, ICN 2015, IFIP Networking 2016]
   - Access Control Dimensions
   - Interest-Based Access Control
   - Revocation and Network Eviction

3 Privacy [CCNC 2016, WPES 2016, IFIP Networking 2017 (x2), ICN 2017, LCN 2017]
   - Privacy Parity
   - Frequency Analysis Attacks

4 Conclusion

Christopher A. Wood

Security and Privacy Challenges in Content-Centric Networks
Motivated content-centricity and CCN

Discussed outstanding security and privacy challenges:

- Access control is an application concern, yet the network plays a role
  - CBAC decouples authentication and authorization for content consumption
  - IBAC couples authentication and authorization at the time of request
  - Revocation requires support from the network

- Network efficiencies pay the price of data privacy
  - Strong privacy requires end-to-end encryption
  - Weak privacy leaks information and is subject to frequency analysis attacks
Current & Future Work

- Compare privacy of IP-based applications to CCN-based applications
  - Is CCN better, worse, or different?
- Expand statistical analysis of the privacy frequency analysis attack
  - How does it fare with different distributions?
- Merge BEAD and secure accounting for a scalable content deletion service
  - Autonomous System (AS) gateways generate “private push interests”
  - Producers send deletion requests to origin AS gateways
- Define a CCN API to help applications transition from IP-based connections to CCN
CCN Related Publications


C. A. Wood and J. Jacob, Characterization of Small Trees Based on their L(2,1)-Span, AKCE International Journal of Graphs and Combinatorics, Volume 12, Issue 1, July 2015, Pages 2631.


Very special thanks to:

- Gene, Ersin, and Marco
- SPROUT colleagues, old and new
- PARC and Apple colleagues
- Family and friends
Questions?