Application-Specific Secure Gathering of Consumer Preferences and Feedback in Information-Centric Networks

Reza Tourani, Satyajayant (Jay) Misra, Travis Mick

Computer Science Department
New Mexico State University
Outline

- Introduction and Motivation
- Problem Definition
- Requirements and Preliminaries
- Feedback Collection and Delivery Approaches
- Conclusions and Future Work
Outline

- Introduction and Motivation
- Problem Definition
- Requirements and Preliminaries
- Feedback Collection and Delivery Approaches
- Conclusions and Future Work
Client mining is widespread.
Benefits of client mining and recommender systems.

Influence on 80% of hours streamed at Netflix (2016)

Approximately 35% increase in Amazon revenue (2013)

50% of LinkedIn job applications and job views by members (2011)
Netflix Communication Flow

Netflix Data Center (www.netflix.com)

Amazon Cloud

Netflix Server

CDN Routing

DRM

Redirecting users

Authentication

Manifest File

Periodic Updates

New user registration

User account billing

Client

CDNs

New Mexico State University, NM
Outline

- Introduction and Motivation
- Problem Definition
- Requirements and Preliminaries
- Feedback Collection and Delivery Approaches
- Conclusions and Future Work
Request flow in ICN multi-level architecture.
Data flow in ICN multi-level architecture.
Pervasive caching eliminates contacting provider for popular content.
How to track client without communication?

Caching undermines gathering of access statistics.
Outline

- Introduction and Motivation
- Problem Definition
- Requirements and Preliminaries
- Feedback Collection and Delivery Approaches
- Conclusions and Future Work
ICN requirements for successful client mining.

- Preserving User Privacy
- Precise Statistics
- Content Provider Independent

Secure feedback collection

New Mexico State University, NM
Content Categorization

**Static Content**
- Generated beforehand
- Generated in advance
- Available publicly
- Require access control
- Cacheable

**Dynamic Content**
- Generated on-demand
- Generated by request
- Available publicly
- Require access control
- Non-cacheable

**Public Content**
- Publicly available
- Generated by request
- Generated in advance
- Cacheable

**Private Content**
- Require access control
- Generated by request
- Generated in advance
- Non-cacheable
Static-Public is the largest content category.

Content Type in North America

- Static-Public: 66%
- Other: 34%
A bigger portion of mobile access traffic is encrypted in comparison to fixed access traffic.

Fixed Access Traffic in North America 2015
- Encrypted: 29%
- Un-Encrypted: 71%

Fixed Access Traffic in North America 2016
- Encrypted: 36%
- Un-Encrypted: 64%

Mobile Access Traffic in North America 2016
- Encrypted: 37%
- Un-Encrypted: 63%


New Mexico State University, NM
Outline

- Introduction and Motivation
- Problem Definition
- Requirements and Preliminaries
- Feedback Collection and Delivery Approaches
- Conclusions and Future Work
Feedback Collection and Delivery

Preference Tracking Mechanisms

Manifest-Free
- Collection by Intermediate routers
- Collection by Clients
- Collection by ISP's Server

Manifest-Based
- Manifest from Provider
- Manifest from ISP's Server

New Mexico State University, NM
Drawbacks of Collection by Intermediate Routers

Collection Event

- Per-Interest
  - Redundant Statistics
  - Coarse-level Statistics
  - Computation Overhead
  - Lack of Client ID

- Per-Hit
  - Coarse-level Statistics
  - Computation Overhead
  - Lack of Client ID
Drawbacks of Collection by Clients

**Approaches**

**Content Partitioning**
- Drawbacks
  - Unknown Partition Size
  - Partition Publication
  - Dependency on Online Server
  - Communication Overhead

**Access Control Enforcement**
- Drawbacks
  - Suitable for Private Content
  - Dependency on Online Server
  - Communication Overhead

New Mexico State University, NM
Collection by the ISP’s Designated Server

**Provider**
- Offload Decryption Key or Content Partition

**ISP's Server**
- Stores Statistics
- Returns Requested Key or Content

**User**
- Request Decryption Key or Content Partition

**Benefits**
- Reduced Latency
- Cache Utilization
- Independent of Provider

**Drawbacks**
- ISP-Provider Interaction
- Inaccurate Statistics

New Mexico State University, NM
Manifest-Based Approaches

Manifest Delivery

Provider (Un-cacheable)

Drawbacks
- Extra Latency
- Provider’s Availability
- Un-cached Content

ISP’s Server (Cacheable)

Drawbacks
- Single Point of Failure
- Network Bottleneck
Outline

- Introduction and Motivation
- Problem Definition
- Requirements and Preliminaries
- Feedback Collection and Delivery Approaches
- Conclusions and Future Work

New Mexico State University, NM
How about evaluation?

- Manifest-based approaches scale much better than the other schemes.
- Communication overhead as means of evaluating efficacy of the approaches – Manifest based approaches scale better.
- Manifest-based approaches introduce fixed amount of overhead per content and the amortized cost will be low.
- There is a theoretical upper bound on the required communication overhead per content:

\[
\text{Overhead} = \text{Distance}_{\text{client - server}} \times \text{Manifestsize}
\]
Conclusions and Future Work

- Direct interaction between client and provider with/without help of routers inaccurate and non-scalable.
- A viable feedback collection mechanism should leverage caching.
- Manifest-based feedback collection approaches are more scalable, especially if it involves infrastructure at the ISP.
- Comprehensive evaluation of manifest based approaches (Provider vs. ISP server) and identify which approach in the other class comes closer.
Thank you!

Email: misra@cs.nmsu.edu

Research funded by the US National Science Foundation and the US Dept. of Defense.