An Integrated System Solution for Secure P2P Content Distribution Based on Network Coding

Heng He\textsuperscript{a}, Ruixuan Li\textsuperscript{a}, Guoqiang Gao\textsuperscript{a}, Zhiyong Xu\textsuperscript{b}, Weijun Xiao\textsuperscript{c}

\textsuperscript{a} Huazhong University of Science and Technology
\textsuperscript{b} Suffolk University
\textsuperscript{c} University of Minnesota

IEEE NAS 2011, Dalian, July 28, 2011
**Introduction**

**Network Coding**
- New paradigm of routing:
  - Packet mixing at intermediate nodes

![Diagram showing traditional store-and-forward and network coding](image)

- **Benefits:**
  - Maximum throughput, robustness to link failure, energy efficiency ...

- **Applications:**
  - Multicast/broadcast, P2P file distribution, P2P streaming, wireless unicast ...
Network coding in P2P content distribution

• Benefits of network coding in P2P content distribution
  Better resilience to peer dynamics
  leading to less downloading time

• Drawbacks
  Network coding is vulnerable to pollution attacks

  The traditional signatures and hashes
  can not protect encoded blocks!
Related work

To thwart pollution attacks in network coding:

- **Corruption detection**
  - References [Infocom 08, 09]
    - expensive for P2P system
    - vulnerable to collusion attack

- **Error correction**
  - References [Koetter, IEEE TIT]
    - not applicable for P2P system

- **Attacker identification**
  - Attacker identification is a more efficient approach in P2P system
  - It has received much less attention

- **Attacker identification**
  - References [Wang, Infocom 2010]
    - significant identification overhead in dynamic P2P network
    - vulnerable to collusion attack
    - low security level
**ISNC**: An Integrated system solution for Secure P2P content distribution based on Network Coding

**Objective**
- Detect corrupted blocks on-the-fly
- Identify malicious peers effectively
- Maintain high throughput
- Be applicable for P2P system

**Contributions**
- The system architecture based on extended uniform bipartite network
- A secure network coding signature scheme
- An identity-based malicious peer identification scheme
Outline

• The system architecture
• A secure network coding signature scheme
• An identity-based malicious peer identification scheme
• Performance evaluation
• Conclusion
• Extend the uniform bipartite network as system topology, achieving high throughput with network coding.

- The above three layers constitute a uniform bipartite network $C_n^k (n=6, k=3)$;
- Peers in the 4th layer and below connect with $k$ peers in the upper layers.
• Utilize linear network coding to encode every group of the file, to reduce coding complexity.

\[ B_1 = \sum_{i=1}^{k} c_i \cdot b_i , \text{ Coding vector } (c_1, c_2, ..., c_k) \]

\[ B_3 = c_1'' B_1 + c_2'' B_2 , \]

Coding vector \((c_1'' c_1 + c_2'' c_1', c_1'' c_2 + c_2'' c_2', ...)\)

A peer reconstructs original blocks when receiving \(k\) independent coded blocks.
A secure network coding signature scheme

• A secure network coding signature scheme, based on homomorphic hash function, achieving high security and applicability of P2P systems.

• Given a file identifier $id_f$, a group identifier $id_g$, $k$ original blocks $b_i$ ($i=1, \ldots, k$), the source computes the signatures as:

  • Algorithm 1
    - Compute the homomorphic hash for each block $b_i = (b_{i1}, \ldots, b_{ir})$ as $\sigma_i = \prod_{j=1}^{r} g_{ij}^{b_{ij}} \mod p$, for $i = 1, \ldots, k$.
    - Compute the signature for the hashes as $\theta = \text{Sign}(SK, (id_f, id_g, \sigma_1, \ldots, \sigma_k))$, where Sign is a standard signature algorithm.
    - Generate the signature of the group $\varphi = (\sigma_1, \ldots, \sigma_k, \theta)$. 
A secure network coding signature scheme

- Peers download the signatures from its upstream peers. The peer checks the validity of block $B = \sum_{i=1}^{k} c_i \cdot b_i$ as:

- Algorithm 2
  - Check the validity of $\varphi$ by standard signature verification algorithm. If $\varphi$ is invalid, algorithm2 aborts. The peer must contact its upstream peers to regain $\varphi$.
  - Compute the homomorphic hash of $B = (B_1, ..., B_r)$ as $\sigma = \prod_{j=1}^{r} g_j^{B_j}$ mod $p$.
  - Compute the hash of $B$ as $\sigma' = \prod_{i=1}^{k} \sigma_i^{c_i}$ mod $p$.
  - If $\sigma = \sigma'$, the receiving peer accepts $B$; otherwise, it discards corrupted $B$. 
An identity-based malicious peer identification scheme

- Blocks not be checked are kept in an insecure window

Computation of homomorphic hashes may be expensive for some peers

- Peers check blocks probabilistically
  - Blocks are checked using batching method

- When detecting corrupted blocks using batching, an identity-based malicious peer identification scheme is triggered

  - Prevent corrupted blocks from propagating
  - Identify malicious peers
An identity-based malicious peer identification scheme

- The scheme is based on the uniform bipartite network topology.

- Every peer maintains a table containing:
  1) the identities of its upstream neighbors that sent blocks inside its insecure window $ID_u$, the time stamps of these blocks $TS_u$
  2) the identities of its downstream neighbors that received blocks encoded with insecure window blocks $ID_d$, the time stamps of the encoded blocks $TS_d$
**An identity-based malicious peer identification scheme**

- **A Bottom-up Approach for Identification Scope Restriction**

  **Step 1:**
  Send a trace message and an alert message (trace message contains $id_g$, $ID_u$, $TS_u$; alert message contains $id_g$, $ID_d$, $TS_d$)

  **Step 2:**
  Send trace messages to those upstream neighbors of A with received $ID_u$.

  **Step 3:**
  Send alert messages to those downstream neighbors of A with received $ID_d$.

  **Step 4:**
  - Check whether A in $ID_d$ and the given time stamps in $TS_d$
  - Send messages like A, or send a halt message to the server

  **Step 5:**
  - Check whether A in $ID_u$ and the given time stamps in $TS_u$
  - Send alert message like A, or the alert message stops at C

  **Step 6:**
  Send combined trace messages if multiple peers send messages with the same upstream neighbors

  **Record the messages**
An identity-based malicious peer identification scheme

• A Bottom-up Approach for Identification Scope Restriction

- the identification scope restricted as the area that trace messages cover
- peers in the scope start to check blocks concurrently after sending messages
An identity-based malicious peer identification scheme

• A Top–down Approach for Malicious Peer Identification

- The server identifies malicious peers from the highest to the lowest layer
- E.g. to check Q, downstream neighbors return results to the server
- The peer doesn’t need to be checked if its upstream neighbor is malicious
Performance evaluation

- Throughput Evaluation (through simulations)
  - network size: 200 peers; file size: 100 blocks; group size: 6 blocks
  - source & relay peers: 6 blocks/round; other peers: 3 blocks/round
  - topology, $C_{12}^6$ extended uniform bipartite network

  - Average finish time of our scheme is 15% less than BitTorrent
  - Our system brings better throughput
Performance evaluation

- Corruption Evaluation
  - malicious peer proportions: 15%, 25%, 35% in 3 circumstances
  - probability of checking: 5%

- The scheme proposed in [Infocom 06]:

<table>
<thead>
<tr>
<th>Pm</th>
<th>15%</th>
<th>25%</th>
<th>35%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of corrupted peers</td>
<td>20%</td>
<td>24%</td>
<td>26%</td>
</tr>
</tbody>
</table>

- All malicious peers are identified by ISNC
- Percentage of corrupted peers is smaller in ISNC
Performance evaluation

- System Overhead
  - signature size: only 0.05% of file size
  - peers download signatures only once, no signature added to transmitted blocks
  - Current existing schemes repeatedly distribute verification information and append signature to every block, which brings significant overheads
Conclusion

• We propose a novel and integrated system solution for secure P2P content distribution based on network coding (ISNC) against pollution attacks.

• ISNC can not only detect corrupted blocks effectively, but also identify all the malicious peers, even when they collude to launch attacks.

• ISNC is especially applicable for P2P content distribution, and can achieve both high security and overall efficiency.
THANK YOU!

Q/A?