Byzantine modification detection for multicast using network coding

T. Ho, B. Leong, R. Koetter, M. Médard, M. Effros, D. R. Karger

Overview

Randomized linear network coding is a robust and efficient approach to multicasting. By including a hash value in each packet that is a simple polynomial function of the data, receivers can detect the presence of arbitrarily modified packets without needing to obtain additional packets, as long as the modified packets have not been designed with complete knowledge of other received packets.

Model and approach

• Length-$n$ vectors of data bits viewed as elements of finite field $F_q$, where $q = 2^n$ [1]

• Each source packet is a sequence of $d$ data symbols $x_1, x_2, ..., x_d$ and $c$ hash symbols $y_1, y_2, ..., y_c$ from $F_q$, calculated using function $\pi(x_1, x_2, ..., x_k) = x_1^2 + x_2^3 + ... + x_k^{k+1}$ as follows:

<table>
<thead>
<tr>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
<th>$s_4$</th>
<th>$s_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
<td>$x_3$</td>
<td>$x_4$</td>
<td>$x_5$</td>
</tr>
<tr>
<td>$y_1$</td>
<td>$y_2$</td>
<td>$y_3$</td>
<td>$y_4$</td>
<td>$y_5$</td>
</tr>
</tbody>
</table>

where $k = \lceil d/c \rceil$ is a parameter trading off overhead against detection probability

• Blocks of $r$ source packets undergo random linear combinations at network nodes; each packet contains a vector of coefficients describing its linear composition [2]

• A receiver with $r$ linearly independent packets can decode and check hash values

Performance and discussion

Suppose a Byzantine attacker supplies $v$ linearly independent modified packets without knowing the contents of $s$ unmodified packets obtained by a receiver.

• The attacker cannot determine which of $q^v$ potential decoding outcomes will be obtained

• At most a fraction $\left( \frac{s+1}{q} \right)$ of these outcomes can have consistent data and hash values, i.e. detection probability is at least $\left( \frac{s+1}{q} \right)$

• E.g. with 2% overhead ($k=50$), code length=7, $s=5$, detection probability $\geq 98.9%$

• with 1% overhead ($k=100$), code length=8, $s=5$, detection probability $\geq 99.0%$

• Detection probability increases with the overhead, code complexity and the number of unmodified packets obtained at the receiver whose contents are unknown to the attacker

No cryptographic functions or bounds on the number/proportion of Byzantine nodes are needed

• Efficient approach for verifying data integrity

• Depending on the application, various responses may be employed upon detection of the Byzantine fault, such as

• collecting more packets from different nodes and testing various sets until a consistent decoding set is found

• using an error correction approach

• employing a more complex Byzantine agreement algorithm

Analysis

• Let $M$ be the matrix whose rows $w_i$ represent the data and hash value of each source packet

• An unmodified packet contains a random linear combination of rows of $M$, together with the vector of coefficients of the combination

• Suppose the receiver tries to decode using $s$ unmodified packets, represented $C_s[M \mid I]$

• $r$-s modified packets, represented by $[C_sM + F \mid C_b]$ where $V$ is an arbitrary matrix

• Consider any fixed $C_s$, $V$, and $r$. Since the receiver decodes only when it has a full rank set of packets, possible values of $C_s$ have rows independent of those of $C_b$

It can be shown that

• for at least $s$ packets, the attacker knows only that the decoded value will be one of $q^{r\text{ rank}(F)}$ possibilities $\left[ m + \sum_{j=1}^{\text{rank}(F)} y_{ij} \in F_q \right]$

• at most $k+1$ out of the $q$ vectors in a set $\{ y + \gamma \mid y \in F_q \}$ where $\gamma = (\gamma_1, \gamma_2, ..., \gamma_k)$ is a fixed length-$k$ vector and $y = (y_1, y_2, ..., y_k)$ a fixed nonzero length-$(k+1)$ vector, can satisfy the property that the last element of the vector equals the hash of the first $k$ elements

References