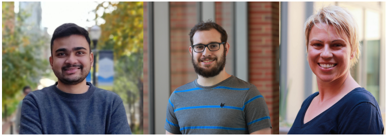
Concentrated Stopping Set Design for Coded Merkle Tree: Improving Security Against Data Availability Attacks in Blockchain Systems

Debarnab Mitra, Lev Tauz, and Lara Dolecek

Electrical and Computer Engineering
University of California, Los Angeles

Non-Volatile Memories Workshop
May 10, 2022



UCLA Samueli School of Engineering

Table of Contents

Blockchain

- Institutional Trust Systems

Blockchain

- Institutional Trust Systems

Blockchain

- Institutional Trust Systems

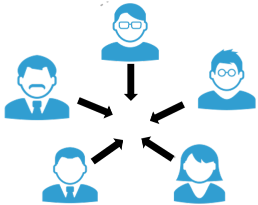
  All parties trust an established institution

Blockchain

- Institutional Trust Systems

  All parties trust an established institution

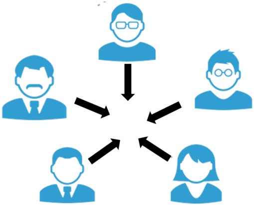
- Distributed/Decentralized Trust Systems



Blockchain

- Institutional Trust Systems
  All parties trust an established institution

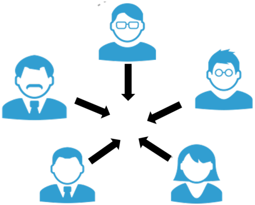
- Distributed/Decentralized Trust Systems
  Multiple parties collaborate on a specific task without parties trusting one another



Blockchain

- Institutional Trust Systems
  - All parties trust an established institution

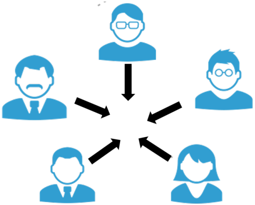
- Distributed/Decentralized Trust Systems
  - Multiple parties collaborate on a specific task without parties trusting one another



Blockchain:

Blockchain

- Institutional Trust Systems
  - All parties trust an established institution



- Distributed/Decentralized Trust Systems
  - Multiple parties collaborate on a specific task without parties trusting one another

Blockchain: Allows for decentralized trust systems

Blockchain

- Institutional Trust Systems

  All parties trust an established institution

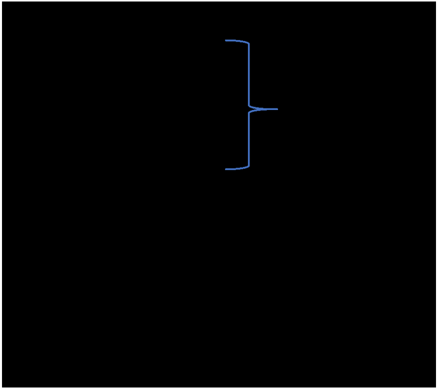
- Distributed/Decentralized Trust Systems

  Multiple parties collaborate on a specific task without parties trusting one another

Blockchain: Allows for decentralized trust systems

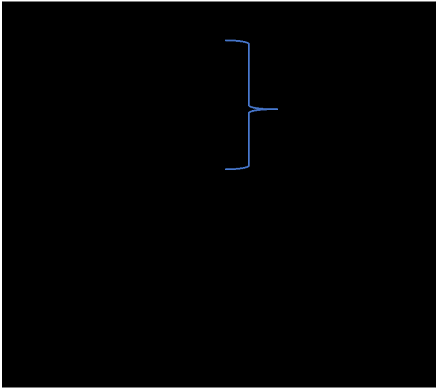
  Main application of Blockchain: Currency and Finance

Blockchain Ledger



- Ledger of transactions

Blockchain Ledger



- Ledger of transactions
- Arranged in the form of blocks

Blockchain Ledger

- Ledger of transactions
- Arranged in the form of blocks
- Stored by a network of nodes

Blockchain Ledger

- Ledger of transactions
- Arranged in the form of blocks
- Stored by a network of nodes
- Nodes store the ledger in their memory

Blockchain Ledger

- Ledger of transactions
- Arranged in the form of blocks
- Stored by a network of nodes
- Nodes store the ledger in their memory
- Operate on the ledger to validate transactions

Blockchain Ledger

- Ledger of transactions
- Arranged in the form of blocks
- Stored by a network of nodes
- Nodes store the ledger in their memory
- Operate on the ledger to validate transactions

NVM technologies like persistent memories: low latency, high reliability

Blockchain Ledger

- Ledger of transactions
- Arranged in the form of blocks
- Stored by a network of nodes
- Nodes store the ledger in their memory
- Operate on the ledger to validate transactions

NVM technologies like persistent memories: low latency, high reliability
Faster blockchains

Central Problem: Prohibitive Storage Overhead

- Each node stores a copy of the ledger in its memory

Central Problem: Prohibitive Storage Overhead

- Each node stores a copy of the ledger in its memory

Signi cant storage overhead

Central Problem: Prohibitive Storage Overhead

- Each node stores a copy of the ledger in its memory

Signi cant storage overhead

- Bitcoin ledger size 300GB[1]
- Ethereum ledger size 650GB[2]

As of 4/28/2022, [1]https://www.blockchain.com/charts/blocks-size
[2]https://etherscan.io/chartsync/chaindefault

Central Problem: Prohibitive Storage Overhead

- Each node stores a copy of the ledger in its memory
- Prohibitive for resource limited nodes

Signi cant storage overhead

- Bitcoin ledger size 300GB[1]
- Ethereum ledger size 650GB[2]

As of 4/28/2022, [1]https://www.blockchain.com/charts/blocks-size
[2]https://etherscan.io/chartsync/chaindefault

Allowing Light Nodes

Light Nodes:

Allowing Light Nodes

Light Nodes:

- Only store block headers
  (total size 1GB for Ethereum)

Allowing Light Nodes

Light Nodes:

- Only store block headers
  (total size 1GB for Ethereum)
- Can verify transaction inclusion in a block

Allowing Light Nodes

Light Nodes:

- Only store block headers
  (total size 1GB for Ethereum)
- Can verify transaction inclusion in a block
- Cannot verify transaction correctness

Allowing Light Nodes

Light Nodes:

- Only store block headers (total size   1GB for Ethereum)
- Can verify transaction inclusion in a block
- Cannot verify transaction correctness ! Rely on honest Full nodes for fraud noti cation

Allowing Light Nodes

Light Nodes:

- Only store block headers
  (total size   1GB for Ethereum)
- Can verify transaction inclusion in a block
- Cannot verify transaction correctness
  !   Rely on honest Full nodes for fraud noti cation
- Full nodes send veri able fraud proofs to the light nodes to reject invalid blocks

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

- Adversary: Provides block to Full node but hides invalid portion

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

- Adversary: Provides block to Full node but hides invalid portion
  Provides header to Light node

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

- Adversary: Provides block to Full node but hides invalid portion
  Provides header to Light node
- Honest Nodes: Cannot verify missing transactions

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

- Adversary: Provides block to Full node but hides invalid portion
  Provides header to Light node
- Honest Nodes: Cannot verify missing transactions No fraud proof

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

- Adversary: Provides block to Full node but hides invalid portion
  Provides header to Light node
- Honest Nodes: Cannot verify missing transactions No fraud proof
- Light Nodes: No fraud proof

Data Availability(DA) Attack

Systems with light nodes and a dishonest majority of full nodes are vulnerable to DA attacks [Al-Bassam '18], [Yu '19]

Adversary creates an invalid block

- Adversary: Provides block to Full node but hides invalid portion
  Provides header to Light node
- Honest Nodes: Cannot verify missing transactions No fraud proof
- Light Nodes: No fraud proof! Accept the header

Ensuring Data Availability

Ensuring Data Availability

Ensuring Data Availability

- Request/sample few random chunks of the block

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

Probability of failure using 2 random samples:

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

Probability of failure
using 2 random samples:

$$1 \quad \frac{1}{16} \quad 1 \quad \frac{1}{15} = 0{:}87$$

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

No coding:

Probability of failure using 2 random samples:

$$1 \quad \frac{1}{16} \quad 1 \quad \frac{1}{15} = 0{:}87$$

Erasure coding:

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

No coding:

Probability of failure using 2 random samples:

$$1 \quad \frac{1}{16} \quad 1 \quad \frac{1}{15} = 0{:}87$$

Erasure coding:

Ensuring Data Availability

- Request/sample few random chunks of the block
- Adversary can hide a small portion

No coding:

Probability of failure using 2 random samples:

$$1 \quad \frac{1}{16} \quad 1 \quad \frac{1}{15} = 0{:}87$$

Erasure coding:

Probability of failure using 2 random samples:

$$1 \quad \frac{17}{32} \quad 1 \quad \frac{17}{31} = 0{:}21$$

Choice of Code Matters

Choice of Code Matters

- Incorrect coding attack:

Choice of Code Matters

- Incorrect coding attack:

  Adversary sends incorrectly coded block to Full Nodes

Choice of Code Matters

- Incorrect coding attack:
  - Adversary sends incorrectly coded block to Full Nodes
  - Honest Full nodes can detect and send incorrect coding proof
  - Incorrect coding proof size: O(sparsity of parity check equations)

Choice of Code Matters

- Incorrect coding attack:
  - Adversary sends incorrectly coded block to Full Nodes
  - Honest Full nodes can detect and send incorrect coding proof
  - Incorrect coding proof size: O(sparsity of parity check equations)
  - MDS codes: proof size = O(block size)

Choice of Code Matters

- Incorrect coding attack:
  - Adversary sends incorrectly coded block to Full Nodes
  - Honest Full nodes can detect and send incorrect coding proof
  - Incorrect coding proof size: O(sparsity of parity check equations)
  - MDS codes: proof size = O(block size)
- Decoding complexity

Choice of Code Matters

- Incorrect coding attack:
  - Adversary sends incorrectly coded block to Full Nodes
  - Honest Full nodes can detect and send incorrect coding proof
  - Incorrect coding proof size: O(sparsity of parity check equations)
  - MDS codes: proof size = O(block size)
- Decoding complexity
- Undecodable ratio

Choice of Code Matters

- Incorrect coding attack:
  - Adversary sends incorrectly coded block to Full Nodes
  - Honest Full nodes can detect and send incorrect coding proof
  - Incorrect coding proof size: O(sparsity of parity check equations)
  - MDS codes: proof size = O(block size)
- Decoding complexity
- Undecodable ratio
  - Probability of Light node failure using s random samples = $(1 \quad)^s$

LDPC Codes: A Strong Contender

LDPC codes:

- Characterized by a sparse parity check matrix

LDPC Codes: A Strong Contender

LDPC codes:

- Characterized by a sparse parity check matrix
- Tanner Graph

LDPC Codes: A Strong Contender

LDPC codes:

- Characterized by a sparse parity check matrix

- Tanner Graph

LDPC codes have been shown to be suitable for this application [Yu' 19

LDPC Codes: A Strong Contender

LDPC codes:

- Characterized by a sparse parity check matrix
- Tanner Graph

LDPC codes have been shown to be suitable for this application [Yu' 19

- Small incorrect coding proof size due to small check node degree

LDPC Codes: A Strong Contender

LDPC codes:

- Characterized by a sparse parity check matrix
- Tanner Graph

LDPC codes have been shown to be suitable for this application [Yu' 19

- Small incorrect coding proof size due to small check node degree
- Linear decoding in terms of the block size using peeling decoder

LDPC Codes: A Strong Contender

LDPC codes:

- Characterized by a sparse parity check matrix
- Tanner Graph

LDPC codes have been shown to be suitable for this application [Yu' 19

- Small incorrect coding proof size due to small check node degree
- Linear decoding in terms of the block size using peeling decoder

  What about the undecodable ratio?

Challenge with LDPC Codes: Small Stopping Sets

- Substructure in the Tanner Graph

Challenge with LDPC Codes: Small Stopping Sets

- Substructure in the Tanner Graph
- If hidden, prevents peeling decoder from decoding the block No fraud proof

Challenge with LDPC Codes: Small Stopping Sets

- Substructure in the Tanner Graph
- If hidden, prevents peeling decoder from decoding the block No fraud proof

Probability of failure using 2 random samples:

$$1 \quad \frac{3}{32} \quad 1 \quad \frac{3}{31} = 0.81$$

Challenge with LDPC Codes: Small Stopping Sets

- Substructure in the Tanner Graph
- If hidden, prevents peeling decoder from decoding the block No fraud proof

Probability of failure using 2 random samples:

$$1 \quad \frac{3}{32} \quad 1 \quad \frac{3}{31} \quad = 0{:}81$$

Our work: Design of specialized LDPC codes with a coupled sampling strategy to achieve a signi cantly lower probability of failure.

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

Relevant in IoT blockchains [Wang '19], Proof-of-Stake blockchains [Daian '1

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

- Relevant in IoT blockchains [Wang '19], Proof-of-Stake blockchains [Daian '1
- Stronger adversaries have been considered in full paper [Mitra '21]

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

- Relevant in IoT blockchains [Wang '19], Proof-of-Stake blockchains [Daian '1
- Stronger adversaries have been considered in full paper [Mitra '21]

Lemma

Of all stopping sets (SSs) of size, when an adversary randomly hides one of them, and light nodes sample all VNs in the set, then

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

- Relevant in IoT blockchains [Wang '19], Proof-of-Stake blockchains [Daian '1
- Stronger adversaries have been considered in full paper [Mitra '21]

Lemma

Of all stopping sets (SSs) of size, when an adversary randomly hides one of them, and light nodes sample all VNs in the set L, then

Probability of failure = 1 fraction of SSs of size touched by L

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

Relevant in IoT blockchains [Wang '19], Proof-of-Stake blockchains [Daian '1

Stronger adversaries have been considered in full paper [Mitra '21]

Lemma

Of all stopping sets (SSs) of size, when an adversary randomly hides one of them, and light nodes sample all VNs in the set L, then

Probability of failure = 1   fraction of SSs of size touched by L

- Selecting a set L of VNs which touches large no. of SSs

Motivation: Not all VNs are equal

In this presentation, we consider an adversary that randomly hides a stopping set of a particular size.

- Relevant in IoT blockchains [Wang '19], Proof-of-Stake blockchains [Daian '1
- Stronger adversaries have been considered in full paper [Mitra '21]

Lemma

Of all stopping sets (SSs) of size, when an adversary randomly hides one of them, and light nodes sample all VNs in the set L, then

Probability of failure = 1 fraction of SSs of size touched by L

- Selecting a set L of VNs which touches large no. of SSs
  ! Prob. of failure#

Concentrated Stopping Set Design

Code Design Idea:

- Concentrate stopping sets to a small section of VNs

Concentrated Stopping Set Design

SS distribution

Code Design Idea:

- Concentrate stopping sets to a small section of VNs

Concentrated Stopping Set Design

SS distribution

Code Design Idea:

- Concentrate stopping sets to a small section of VNs
- Greedily Sample this small section of VNs

How to Concentrate Stopping Sets?

- When there are no degree-1 VNs, stopping sets are either cycles or interconnection of cycles [Tian '03]

How to Concentrate Stopping Sets?

- When there are no degree-1 VNs, stopping sets are either cycles or interconnection of cycles [Tian '03]
- Concentrating cycles $\Longrightarrow$ Concentrating stopping sets

How to Concentrate Stopping Sets?

- When there are no degree-1 VNs, stopping sets are either cycles or interconnection of cycles [Tian '03]
- Concentrating cycles $\Longrightarrow$ Concentrating stopping sets
  - Directly concentrating stopping sets during code construction incurs hug complexity

How to Concentrate Stopping Sets?

- When there are no degree-1 VNs, stopping sets are either cycles or interconnection of cycles [Tian '03]
- Concentrating cycles $\Rightarrow$ Concentrating stopping sets
  - Directly concentrating stopping sets during code construction incurs huge complexity

How to design codes with concentrated cycles?

How to Concentrate Stopping Sets?

- When there are no degree-1 VNs, stopping sets are either cycles or interconnection of cycles [Tian '03]
- Concentrating cycles $\Rightarrow$ Concentrating stopping sets
  Directly concentrating stopping sets during code construction incurs huge complexity

How to design codes with concentrated cycles?
We do so by modifying the well-known Progressive Edge Growth (PEG) algorithm

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VN v_j
  Expand Tanner Graph in a BFS fashion

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VNv_j
  Expand Tanner Graph in a BFS fashion
  If 9 CNs not connected tov_j

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If 9 CNs not connected to v_j
    Select a CN with min degree not connected to v_j

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If 9 CNs not connected to v_j
    Select a CN with min degree not connected to v_j
  Else

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If 9 CNs not connected to v_j
    Select a CN with min degree not connected to v_j
  Else
    Find CNs most distant to v_j
    Select one with minimum degree

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If 9 CNs not connected to v_j
    Select a CN with min degree not connected to v_j
  Else
    Find CNs most distant to v_j
    Select one with minimum degree
     New cycles created

PEG Algorithm

- Constructs a Tanner Graph in an edge by edge manner [Xiao '05]

For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If 9 CNs not connected to v_j
    Select a CN with min degree not connected to v_j
  Else
    Find CNs most distant to v_j
    Select one with minimum degree
      New cycles created

We modify the CN selection criteria in green to concentrate cycles

Using Entropy to Concentrate Cycles

For distribution $p = (p_1; p_2; \dots; p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

Using Entropy to Concentrate Cycles

For distribution $p = (p_1; p_2; \ldots; p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy

High Entropy

Using Entropy to Concentrate Cycles

For distribution $p = (p_1, p_2, \ldots, p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy Low Entropy

Using Entropy to Concentrate Cycles

For distribution $p = (p_1; p_2; \dots; p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy Low Entropy

We want the cycle distributions to be concentrated

Using Entropy to Concentrate Cycles

For distribution $p = (p_1, p_2, \ldots, p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy Low Entropy

We want the cycle distributions to be concentrated

→ Select CNs such that the entropy of the cycle distribution is minimized

Using Entropy to Concentrate Cycles

For distribution $p = (p_1; p_2; \ldots; p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy

Low Entropy

EC (Entropy Constrained)-PEG Algorithm
For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If $\exists$ CNs not connected to v_j
    select a CN with min degree not connected to v_j
  Else New cycles created

We want the cycle distributions to be concentrated

$\rightarrow$ Select CNs such that the entropy of the cycle distribution is minimized

Using Entropy to Concentrate Cycles

For distribution $p = (p_1, p_2, \ldots, p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy   Low Entropy

EC (Entropy Constrained)-PEG Algorithm

For each VN v_j

Expand Tanner Graph in a BFS fashion

If $\exists$ CNs not connected to v_j

select a CN with min degree not connected to v_j

Else New cycles created

Find CNs most distant to v_j

We want the cycle distributions to be concentrated

$\rightarrow$ Select CNs such that the entropy of the cycle distribution is minimized

Using Entropy to Concentrate Cycles

For distribution $p = (p_1; p_2; \dots; p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy Low Entropy

EC (Entropy Constrained)-PEG Algorithm
For each VN v_j
Expand Tanner Graph in a BFS fashion
If $\exists$ CNs not connected to v_j
select a CN with min degree not connected to v_j
Else New cycles created
Find CNs most distant to v_j
Select CN that results in minimum entropy of resultant cycle distribution

We want the cycle distributions to be concentrated
$\rightarrow$ Select CNs such that the entropy of the cycle distribution is minimized

Using Entropy to Concentrate Cycles

For distribution $p = (p_1, p_2, \ldots, p_n)$, Entropy $H(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$

- Uniform distributions have high entropy
- Concentrated distributions have low entropy

High Entropy Low Entropy

EC (Entropy Constrained)-PEG Algorithm
For each VN v_j
  Expand Tanner Graph in a BFS fashion
  If $\exists$ CNs not connected to v_j
    select a CN with min degree not connected to v_j
  Else New cycles created
    Find CNs most distant to v_j
    Select CN that results in minimum entropy of resultant cycle distribution
    Update cycle distribution

We want the cycle distributions to be concentrated

$\rightarrow$ Select CNs such that the entropy of the cycle distribution is minimized

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

$$\text{VNs}\,(v_1; v_2; :::; v_n)$$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; \ldots; v_n)$

- $^{(g)}_{i}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; \ldots; v_n)$

- $^{(g)}_{i}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1, v_2, \ldots, v_n)$

- ${}_i^{(g)}$:= No. of cycles of length g that v_i is a part of, $g = 4, 6, 8$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; :::; v_n)$

- $^{(g)}_{i}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; \ldots; v_n)$

- $_i^{(g)}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; :::; v_n)$

- ${}^{(g)}_{i}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$
-

$$
\begin{aligned}
{}^{(6)}_{1} &= {}^{(6)}_{1} + 1 \\
{}^{(6)}_{2} &= {}^{(6)}_{2} + 1 \\
{}^{(6)}_{6} &= {}^{(6)}_{6} + 1
\end{aligned}
$$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; :::; v_n)$

- $^{(g)}_{i}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$
-

$$
\begin{aligned}
^{(6)}_{1} &= {}^{(6)}_{1} + 1 \\
^{(6)}_{2} &= {}^{(6)}_{2} + 1 \\
^{(6)}_{7} &= {}^{(6)}_{7} + 1
\end{aligned}
$$

EC-PEG Algorithm

- Whenever a new edge, that creates cycles, is added to the Tanner Graph, we update the cycle counts of each VN

VNs $(v_1; v_2; :::; v_n)$

- $^{(g)}_{i}$:= No. of cycles of length g that v_i is a part of, $g = 4; 6; 8$
-

$$^{(6)}_{1} = \ ^{(6)}_{1} + 1$$
$$^{(6)}_{3} = \ ^{(6)}_{3} + 1$$
$$^{(6)}_{6} = \ ^{(6)}_{6} + 1$$

EC-PEG Algorithm: CN Selection Procedure

Candidate CNs : c_8, c_9, c_{10}

EC-PEG Algorithm: CN Selection Procedure

Candidate CNs : c_8, c_9, c_{10}

- For each CN candidate, calculate the resultant VN cycle counts

EC-PEG Algorithm: CN Selection Procedure

Candidate CNs : c_8, c_9, c_{10}

- For each CN candidate, calculate the resultant VN cycle counts
- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$

EC-PEG Algorithm: CN Selection Procedure

Candidate CNs : c_8, c_9, c_{10}

- For each CN candidate, calculate the resultant VN cycle counts
- $({}_1^{(4)}; \dots; {}_n^{(4)}); ({}_1^{(6)}; \dots; {}_n^{(6)}); ({}_1^{(8)}; \dots; {}_n^{(8)})$
- $({}_1^{(4)}; \dots; {}_n^{(4)}); ({}_1^{(6)}; \dots; {}_n^{(6)}); ({}_1^{(8)}; \dots; {}_n^{(8)})$

EC-PEG Algorithm: CN Selection Procedure

Candidate CNs : c_8, c_9, c_{10}

- For each CN candidate, calculate the resultant VN cycle counts
- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$
- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$
- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$

EC-PEG Algorithm: CN Selection Procedure

- $({}^{(4)}_{1};\ldots;{}^{(4)}_{n});({}^{(6)}_{1};\ldots;{}^{(6)}_{n});({}^{(8)}_{1};\ldots;{}^{(8)}_{n})$
- $({}^{(4)}_{1};\ldots;{}^{(4)}_{n});({}^{(6)}_{1};\ldots;{}^{(6)}_{n});({}^{(8)}_{1};\ldots;{}^{(8)}_{n})$
- $({}^{(4)}_{1};\ldots;{}^{(4)}_{n});({}^{(6)}_{1};\ldots;{}^{(6)}_{n});({}^{(8)}_{1};\ldots;{}^{(8)}_{n})$

EC-PEG Algorithm: CN Selection Procedure

- $(\ _1^{(4)}; \ldots; \ _n^{(4)}); (\ _1^{(6)}; \ldots; \ _n^{(6)}); (\ _1^{(8)}; \ldots; \ _n^{(8)})$
- $(\ _1^{(4)}; \ldots; \ _n^{(4)}); (\ _1^{(6)}; \ldots; \ _n^{(6)}); (\ _1^{(8)}; \ldots; \ _n^{(8)})$
- $(\ _1^{(4)}; \ldots; \ _n^{(4)}); (\ _1^{(6)}; \ldots; \ _n^{(6)}); (\ _1^{(8)}; \ldots; \ _n^{(8)})$

$\underbrace{(\ _1^{(g)}; \ldots; \ _n^{(g)})}_{\text{cycle counts}}$

EC-PEG Algorithm: CN Selection Procedure

- $({}^{(4)}_{1}, \dots, {}^{(4)}_{n}); ({}^{(6)}_{1}, \dots, {}^{(6)}_{n}); ({}^{(8)}_{1}, \dots, {}^{(8)}_{n})$
- $({}^{(4)}_{1}, \dots, {}^{(4)}_{n}); ({}^{(6)}_{1}, \dots, {}^{(6)}_{n}); ({}^{(8)}_{1}, \dots, {}^{(8)}_{n})$
- $({}^{(4)}_{1}, \dots, {}^{(4)}_{n}); ({}^{(6)}_{1}, \dots, {}^{(6)}_{n}); ({}^{(8)}_{1}, \dots, {}^{(8)}_{n})$

$$\underbrace{({}^{(g)}_{1}, \dots, {}^{(g)}_{n})}_{\text{cycle counts}} \rightarrow \underbrace{\left(\frac{{}^{(g)}_{1}}{\sum_{i=1}^{n} {}^{(g)}_{i}}, \dots, \frac{{}^{(g)}_{n}}{\sum_{i=1}^{n} {}^{(g)}_{i}} \right) := {}^{(g)}}_{\text{normalized counts}}$$

EC-PEG Algorithm: CN Selection Procedure

- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$
- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$
- $({}_1^{(4)}; \ldots; {}_n^{(4)}); ({}_1^{(6)}; \ldots; {}_n^{(6)}); ({}_1^{(8)}; \ldots; {}_n^{(8)})$

$$\underbrace{({}_1^{(g)}; \ldots; {}_n^{(g)})}_{\text{cycle counts}} \; ! \; \underbrace{\left(\frac{{}_1^{(g)}}{\sum_{i=1}^{n} {}_i^{(g)}}; \ldots; \frac{{}_n^{(g)}}{\sum_{i=1}^{n} {}_i^{(g)}}\right) := {}^{(g)}}_{\text{normalized counts}} \; ! \; \underbrace{H\left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3}\right)}_{\text{entropy of combined counts}}$$

EC-PEG Algorithm: CN Selection Procedure

- $({}^{(4)}_{1} ; : : : ; {}^{(4)}_{n}); ({}^{(6)}_{1} ; : : : ; {}^{(6)}_{n}); ({}^{(8)}_{1} ; : : : ; {}^{(8)}_{n}) ! \; H \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$
- $({}^{(4)}_{1} ; : : : ; {}^{(4)}_{n}); ({}^{(6)}_{1} ; : : : ; {}^{(6)}_{n}); ({}^{(8)}_{1} ; : : : ; {}^{(8)}_{n})$
- $({}^{(4)}_{1} ; : : : ; {}^{(4)}_{n}); ({}^{(6)}_{1} ; : : : ; {}^{(6)}_{n}); ({}^{(8)}_{1} ; : : : ; {}^{(8)}_{n})$

$$\underbrace{({}^{(g)}_{1} ; : : : ; {}^{(g)}_{n})}_{\text{cycle counts}} \; ! \; \underbrace{\left(\frac{{}^{(g)}_{1}}{\sum_{i=1}^{n} {}^{(g)}_{i}} ; : : : ; \frac{{}^{(g)}_{n}}{\sum_{i=1}^{n} {}^{(g)}_{i}} \right) :=}_{\text{normalized counts}} \; {}^{(g)} \; ! \; \underbrace{H \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)}_{\text{entropy of combined counts}}$$

EC-PEG Algorithm: CN Selection Procedure

- $({}_1^{(4)}, \ldots, {}_n^{(4)}); ({}_1^{(6)}, \ldots, {}_n^{(6)}); ({}_1^{(8)}, \ldots, {}_n^{(8)})\,!\; H\left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3}\right)$
- $({}_1^{(4)}, \ldots, {}_n^{(4)}); ({}_1^{(6)}, \ldots, {}_n^{(6)}); ({}_1^{(8)}, \ldots, {}_n^{(8)})\,!\; H\left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3}\right)$
- $({}_1^{(4)}, \ldots, {}_n^{(4)}); ({}_1^{(6)}, \ldots, {}_n^{(6)}); ({}_1^{(8)}, \ldots, {}_n^{(8)})$

$$\underbrace{({}_1^{(g)}, \ldots, {}_n^{(g)})}_{\text{cycle counts}} \;!\; \underbrace{\left(\frac{{}_1^{(g)}}{\sum_{i=1}^{n} {}_i^{(g)}}, \ldots, \frac{{}_n^{(g)}}{\sum_{i=1}^{n} {}_i^{(g)}}\right) := {}^{(g)}}_{\text{normalized counts}} \;!\; \underbrace{H\left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3}\right)}_{\text{entropy of combined counts}}$$

EC-PEG Algorithm: CN Selection Procedure

- $({}_{1}^{(4)} ; : : : ; {}_{n}^{(4)}) ; ({}_{1}^{(6)} ; : : : ; {}_{n}^{(6)}) ; ({}_{1}^{(8)} ; : : : ; {}_{n}^{(8)}) ! \ H \ (\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3})$
- $({}_{1}^{(4)} ; : : : ; {}_{n}^{(4)}) ; ({}_{1}^{(6)} ; : : : ; {}_{n}^{(6)}) ; ({}_{1}^{(8)} ; : : : ; {}_{n}^{(8)}) ! \ H \ (\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3})$
- $({}_{1}^{(4)} ; : : : ; {}_{n}^{(4)}) ; ({}_{1}^{(6)} ; : : : ; {}_{n}^{(6)}) ; ({}_{1}^{(8)} ; : : : ; {}_{n}^{(8)}) ! \ H \ (\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3})$

$$\underbrace{({}_{1}^{(g)} ; : : : ; {}_{n}^{(g)})}_{\text{cycle counts}} \ ! \ \underbrace{\left(\frac{{}_{1}^{(g)}}{\sum_{i=1}^{n} {}_{i}^{(g)}} ; : : : ; \frac{{}_{n}^{(g)}}{\sum_{i=1}^{n} {}_{i}^{(g)}} \right) := {}^{(g)}}_{\text{normalized counts}} \ ! \ \underbrace{H\left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)}_{\text{entropy of combined counts}}$$

EC-PEG Algorithm: CN Selection Procedure

- $({}_1^{(4)} ; \ldots ; {}_n^{(4)}); ({}_1^{(6)} ; \ldots ; {}_n^{(6)}); ({}_1^{(8)} ; \ldots ; {}_n^{(8)}) ! \; H \; \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$
- $({}_1^{(4)} ; \ldots ; {}_n^{(4)}); ({}_1^{(6)} ; \ldots ; {}_n^{(6)}); ({}_1^{(8)} ; \ldots ; {}_n^{(8)}) ! \; H \; \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$
- $({}_1^{(4)} ; \ldots ; {}_n^{(4)}); ({}_1^{(6)} ; \ldots ; {}_n^{(6)}); ({}_1^{(8)} ; \ldots ; {}_n^{(8)}) ! \; H \; \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$

CN selection procedure:

EC-PEG Algorithm: CN Selection Procedure

- $\left({}^{(4)}_{1} ;:::; {}^{(4)}_{n} \right); \left({}^{(6)}_{1} ;:::; {}^{(6)}_{n} \right); \left({}^{(8)}_{1} ;:::; {}^{(8)}_{n} \right)! \, H \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$
- $\left({}^{(4)}_{1} ;:::; {}^{(4)}_{n} \right); \left({}^{(6)}_{1} ;:::; {}^{(6)}_{n} \right); \left({}^{(8)}_{1} ;:::; {}^{(8)}_{n} \right)! \, H \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$
- $\left({}^{(4)}_{1} ;:::; {}^{(4)}_{n} \right); \left({}^{(6)}_{1} ;:::; {}^{(6)}_{n} \right); \left({}^{(8)}_{1} ;:::; {}^{(8)}_{n} \right)! \, H \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$

CN selection procedure:

Select CN that results in minimum $H \left(\frac{{}^{(4)} + {}^{(6)} + {}^{(8)}}{3} \right)$

EC-PEG Algorithm: CN Selection Procedure

- $({}_1^{(4)};:::;{}_n^{(4)});({}_1^{(6)};:::;{}_n^{(6)});({}_1^{(8)};:::;{}_n^{(8)})!\,H\left(\frac{{}^{(4)}+{}^{(6)}+{}^{(8)}}{3}\right)$
- $({}_1^{(4)};:::;{}_n^{(4)});({}_1^{(6)};:::;{}_n^{(6)});({}_1^{(8)};:::;{}_n^{(8)})!\,H\left(\frac{{}^{(4)}+{}^{(6)}+{}^{(8)}}{3}\right)$
- $({}_1^{(4)};:::;{}_n^{(4)});({}_1^{(6)};:::;{}_n^{(6)});({}_1^{(8)};:::;{}_n^{(8)})!\,H\left(\frac{{}^{(4)}+{}^{(6)}+{}^{(8)}}{3}\right)$

CN selection procedure:
Select CN that results in minimum $H\left(\frac{{}^{(4)}+{}^{(6)}+{}^{(8)}}{3}\right)$

Note:

- Minimizing the entropy of joint cycle counts ensures that all cycle distributions are concentrated towards the same set of VNs

Sampling Strategy

- Our sampling strategy greedily samples VNs that are part of a large number of cycles

```
g = smallest cycle length in Tanner Graph G
While sample set size < s
    v = VN that is part of largest no. of
    cycles of length g in G
    sample set = sample set ∪ v
    remove v and all incident edges from G
```

Sampling Strategy

- Our sampling strategy greedily samples VNs that are part of a large number of cycles

```
g = smallest cycle length in Tanner Graph G
While sample set size < s
    v = VN that is part of largest no. of
    cycles of length g in G
    sample set = sample set ∪ v
    remove v and all incident edges from G
  If ∄ cycles of length g in G
      g = g + 2
```

Simulation Results

- Code parameters: Code length $= 100$, VN degree $= 4$, Rate $= \frac{1}{2}$, girth $= 6$.

Simulation Results

- Code parameters: Code length $= 100$, VN degree $= 4$, Rate $= \frac{1}{2}$, girth $= 6$.

Fraction of cycles touched

VN index

- VN indices arranged in decreasing order of cycle-6 fractions

Simulation Results

- Code parameters: Code length = 100, VN degree = 4, Rate = $\frac{1}{2}$, girth = 6.

Fraction of cycles touched

VN index

- VN indices arranged in decreasing order of cycle 6 fractions
- Cycle 6 and cycle 8 concentrated towards same set of VNs

Simulation Results

Fraction of SSs of size 11, 12 touched by di erent VNs

Simulation Results

Fraction of SSs of size 11, 12 touched by di erent VNs

SSs of size 11

fraction of SSs touched

VN index

- VN indices arranged in decreasing order of cycle 6 fractions

Simulation Results

Fraction of SSs of size 11, 12 touched by di erent VNs

SSs of size 11

fraction of SSs touched

VN index

SSs of size 12

fraction of SSs touched

VN index

- VN indices arranged in decreasing order of cycle 6 fractions

Simulation Results

Fraction of SSs of size 11, 12 touched by different VNs

SSs of size 11

fraction of SSs touched

VN index

SSs of size 12

fraction of SSs touched

VN index

- VN indices arranged in decreasing order of cycle 6 fractions
- SSs are concentrated towards the same set of VNs as the cycles

Simulation Results

Probability of failure for a stopping set of size

Simulation Results

Probability of failure for a stopping set of size

RS: Random Sampling

Prob. of failure

No of Sampless

Simulation Results

Probability of failure for a stopping set of size

RS: Random Sampling

Prob. of failure

No of Sampless

Simulation Results

Probability of failure for a stopping set of size

RS: Random Sampling
GS: Greedy Sampling

Prob. of failure

No of Sampless

Simulation Results

Probability of failure for a stopping set of size

RS: Random Sampling
GS: Greedy Sampling

Prob. of failure

No of Sampless

Simulation Results

Probability of failure for a stopping set of size

RS: Random Sampling
GS: Greedy Sampling

Prob. of failure

No of Sampless

- Concentrated LDPC codes with Greedy sampling improve the probability of failure

Simulation Results

Probability of failure for a stopping set of size

RS: Random Sampling
GS: Greedy Sampling

Prob. of failure

No of Sampless

- Concentrated LDPC codes with Greedy sampling improve the probability of failure

Note that the probability of failure depends on the fraction of stopping sets touched (by greedy sampling) and not the actual number.

Incorrect Coding Proof Size

- Depends on the maximum check node degree

Rate	Code length	VN degree	Ensemble [Yu '19]	PEG	EC-PEG
$\frac{1}{2}$	100	4	16	9	11
	200	4	16	9	15
$\frac{1}{4}$	100	4	8	7	10
	200	4	8	6	9

Table: Maximum CN degree for di erent codes.

Incorrect Coding Proof Size

- Depends on the maximum check node degree

Rate	Code length	VN degree	Ensemble [Yu '19]	PEG	EC-PEG
$\frac{1}{2}$	100	4	16	9	11
	200	4	16	9	15
$\frac{1}{4}$	100	4	8	7	10
	200	4	8	6	9

Table: Maximum CN degree for di erent codes.

- Concentrated LDPC codes do not sacri ce on the incorrect coding proof size

Conclusion and Extentions

- Summary:
  We provided a specialized code construction technique to concentrat
  stopping sets in LDPC codes

Conclusion and Extentions

- Summary:
  - We provided a specialized code construction technique to concentrat stopping sets in LDPC codes
  - Coupled with a greedy sampling strategy, concentrated LDPC codes reduce the probability of light node failure compared to earlier approaches

Conclusion and Extentions

- Summary:
  - We provided a specialized code construction technique to concentrat stopping sets in LDPC codes
  - Coupled with a greedy sampling strategy, concentrated LDPC codes reduce the probability of light node failure compared to earlier approaches
- Extensions (Mitra '21):
  - Considered stronger adversary models that can selectively pick a stopping set that has a lower probability of being sampled to hide instead of randomly

Conclusion and Extentions

- Summary:
    - We provided a specialized code construction technique to concentrat stopping sets in LDPC codes
    - Coupled with a greedy sampling strategy, concentrated LDPC codes reduce the probability of light node failure compared to earlier approaches
- Extensions (Mitra '21):
    - Considered stronger adversary models that can selectively pick a stopping set that has a lower probability of being sampled to hide instead of randomly
    - Provided optimal sampling strategies and associated coupled LDPC code construction to improve the security against such strong adversaries for a given sample complexity

References

- (Mitra '20) D. Mitra, L. Tauz, and L. Dolecek, \Concentrated Stopping Set Design for Coded Merkle Tree: Improving Security Against Data Availability Attacks in Blockchain Systems",in Proc. of IEEE Information Theory Workshop (ITW) , 2020.
- (Mitra '21) D. Mitra, L. Tauz, and L. Dolecek, \Overcoming Data Availability Attacks in Blockchain Systems: LDPC Code Design for Coded Merkle Tree", 2021, submitted to IEEE Transactions on Communications (available at: https://arxiv.org/abs/2108.13332)
- (Al-Bassam '18) M. Al-Bassam,et al., \Fraud and data availability proofs: Detecting invalid blocks in light clients," International Conference on Financial Cryptography and Data Security, Springer, Cham, 2021.
- (Yu '19) M. Yu, et al., \Coded Merkle Tree: Solving Data Availability Attacks in Blockchains," International Conference on Financial Cryptography and Data Security, Springer, Cham, 2020.

References

- (Wang ' 19) X. Wang, *et al.*, \Survey on blockchain for internet of things," *Computer Commununications*, vol. 136, pp. 10-29, 2019.
- (Daian '19) P. Daian, *et al.*, \Snow white: Robustly recon gurable consensus and applications to provably secure proof of stake," *Financial Cryptography*, 2019.
- (Xiao '05) X.Y. Hu, *et al.*, \Regular and irregular progressive edge-growth tanner graphs," *IEEE Transactions of Information Theory*, vol. 51, no. 1, 2005.
- (Tian '03) T. Tian, *et al.*, \Construction of irregular LDPC codes with low error oors," *in Proc. IEEE International Conference on Communications*, 2003.