オルトプラスラボ Tech セミナー 2016 Fall
2016年 9月 30日 (金)

分散システムとして見たブロックチェーン

首藤 一幸
東京工業大学

Tokyo Tech

首藤 一幸 (42)

しゅどう　かずゆき

低レイヤーマン

1996　早稲田大学 修士課程

1998　早稲田大学 博士課程

2001　産総研　**国研**

2006　ウタゴエ(株)

スタートアップ

2008/12 東工大　**大学**

2009/ 5 未踏 PM

魔法のようなソフト

Java スレッド移送システム MOBA

Java Just-in-Time コンパイラ shuJIT

17,000ダウンロード, 商用実績

P2P の基盤ソフト Overlay Weaver

26,000ダウンロード, 15ヶ国

41ヶ国 673台以上で動作 (データベース)

Overlay Weaver

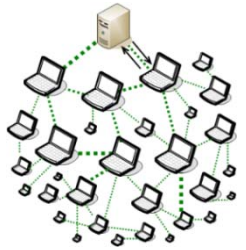
P2P ライブ配信ソフト UG Live

未踏スパクリ × 2人, 商用化, 1万数千人に同時配信

大規模分散システム

書籍 Binary Hacks

著者5人, 1万数千部



P2P のアルゴリズム, 2009 ～

構造化オーバレイ / DHT の統一フレームワーク

分散データベース, 2009 ～

読み書き性能両立, Causal consistency, NVRAM / SCM

分散システムのシミュレーション, 2011 ～

1億ノード / 10台, 既存手法の20倍の性能, Apache Spark 上

少し落ち着いて ブロックチェーンを見極めよう

- 誤解されている？

  ＝ Fintech ？

  ＝ 低コストトランザクション技術 ？

  ＝ 分散データベース ？

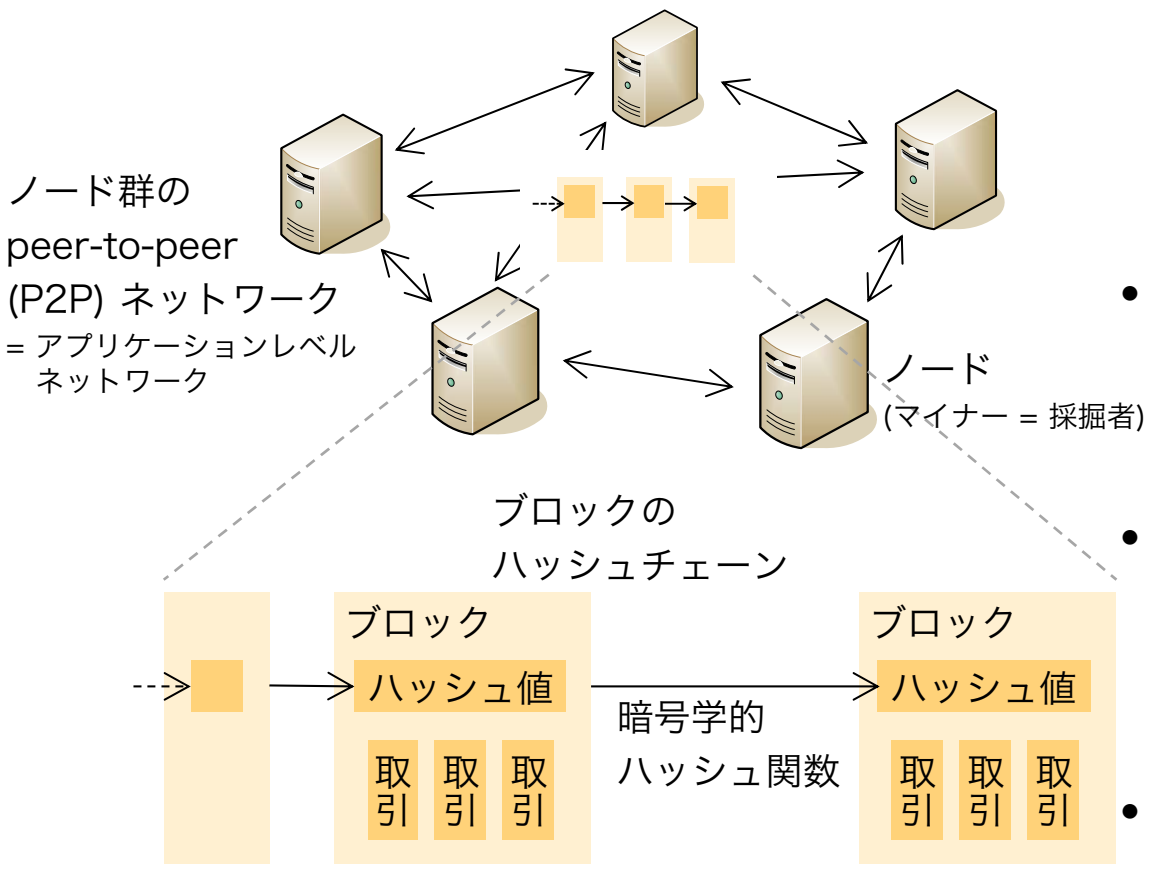
  ＝ 既存データベース管理システム (DBMS) の代替 ？

当セミナー
ウェブページより

- 無理やり使おうとしてない？
  - 「ブロックチェーン」言いたいだけ
  - ハンマーを持つ人にはすべてが釘に見える
    - Abraham Harold Maslow (1908-1970)

Me, too.

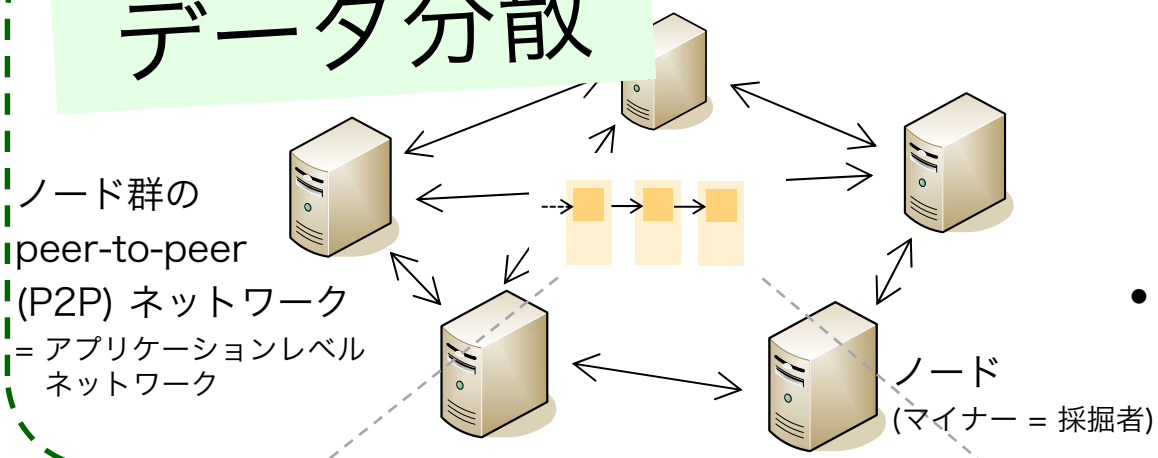
分散システムとして見たブロックチェーン



- ノード群がアプリケーションレベルのネットワークを作る。
  - ノード：Bitcoin で言うマイナー = 採掘者
  - ノード数は数千 ～ 万？
  - 各ノードは、かなり多くの他のノードを把握する。NEM では少なくとも 1,000。
- 基本的に、全ノードがただ１つのハッシュチェーンの全体を持つ。
  - 約10分に1つ (← Bitcoin の場合) ブロックが作られ、ネットワーク全体に行き渡る。
- 各ノードはブロック生成の動機を持つ。ただし、生成の機会は稀少。
  - 動機：貨幣の獲得
  - 機会稀少：POW (Bitcoin 他) / POS / POI (NEM)
- チェーンが分岐している場合、各ノードは最長のチェーンを「確定」と見なす。
  - 逆転の可能性はいつまでも残り、100% の確定はない。

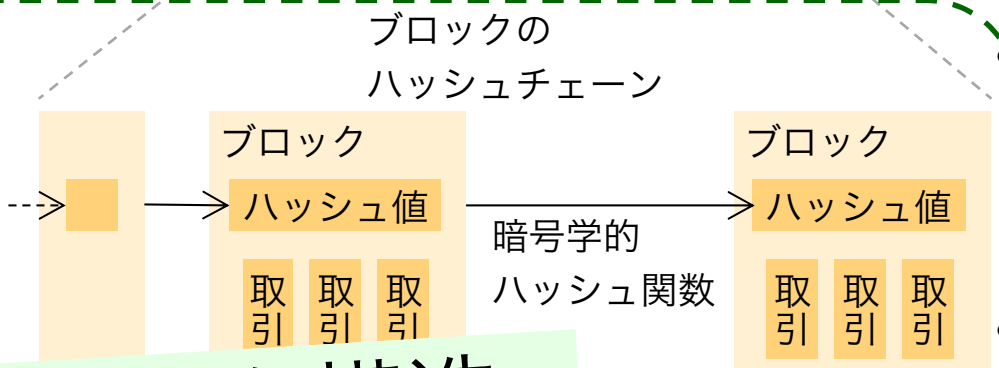
分散システムとして見たブロックチェーン

データ分散



- ノード群がアプリケーションレベルのネットワークを作る。
  - ノード：Bitcoin で言うマイナー = 採掘者
  - ノード数は数千 ～ 万？
  - 各ノードは、かなり多くの他のノードを把握する。NEM では少なくとも 1,000。
- 基本的に、全ノードがただ１つのハッシュチェーンの全体を持つ。
  - 約10分に1つ (← Bitcoin の場合) ブロックが作られ、ネットワーク全体に行き渡る。

データ構造



- 各ノードはブロック生成の動機を持つ。ただし、生成の機会は稀少。
  - 動機：貨幣の獲得
  - 機会稀少：POW (Bitcoin 他) / POS / POI (NEM)
- チェーンが分岐している場合、各ノードは最長のチェーンを「確定」と見なす
  - 逆転の可能性はいつ確定はない。

確定

分散データベースとして見たブロックチェーン

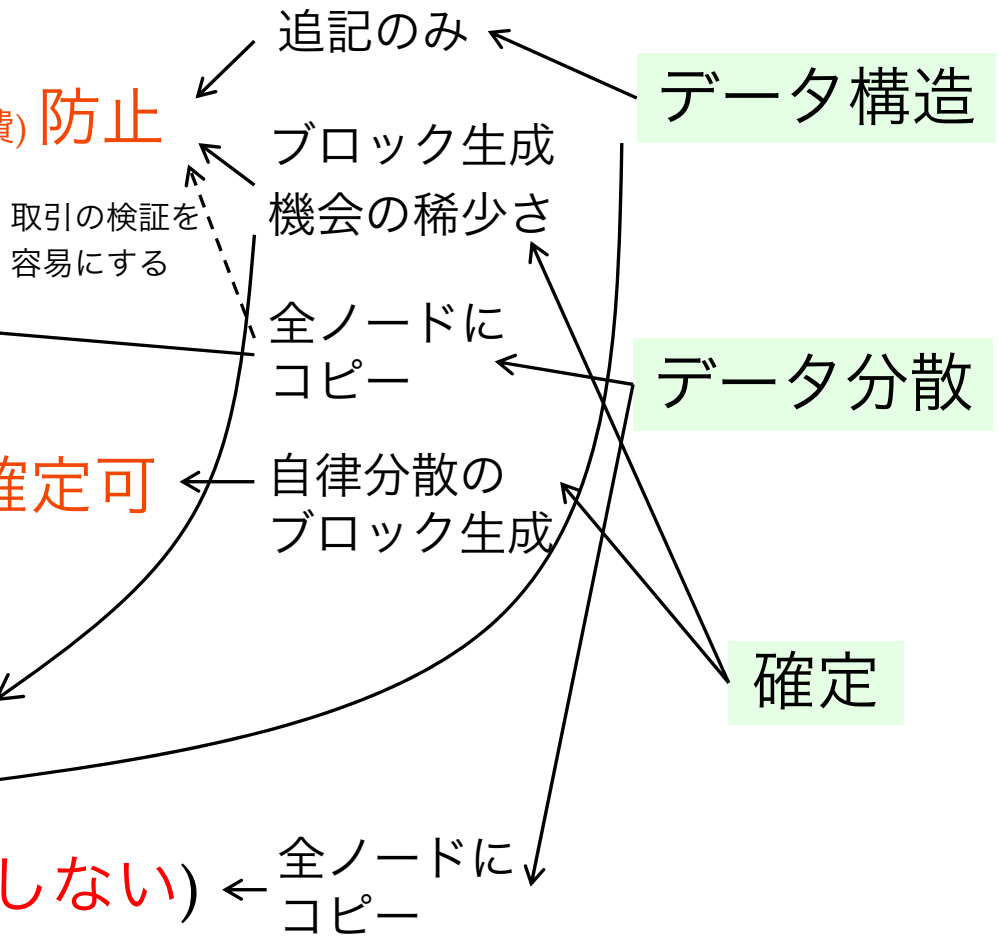
	(Bitcoin) ブロックチェーン	分散 RDB	NoSQL
データ構造	取引を含む**ブロックのハッシュチェーン**	2次元の表	ただのkey-valueペア, 高級な表, …
データ分散	非構造化P2P上での**全ノードへのコピー**	レプリケーション, シャーディング, …	担当ノード決め + いくつかの複製 or coding
確定	**ブロック生成の動機づけ + 機会が稀少**	トランザクション w/ two-phase commit (2PC), …	複製群への伝播 → 結果整合性, …

- 実は広いデザイン空間
  - 「ブロックチェーン」の条件って？

利点と難点、その由来

(Bitcoin) ブロックチェーンの

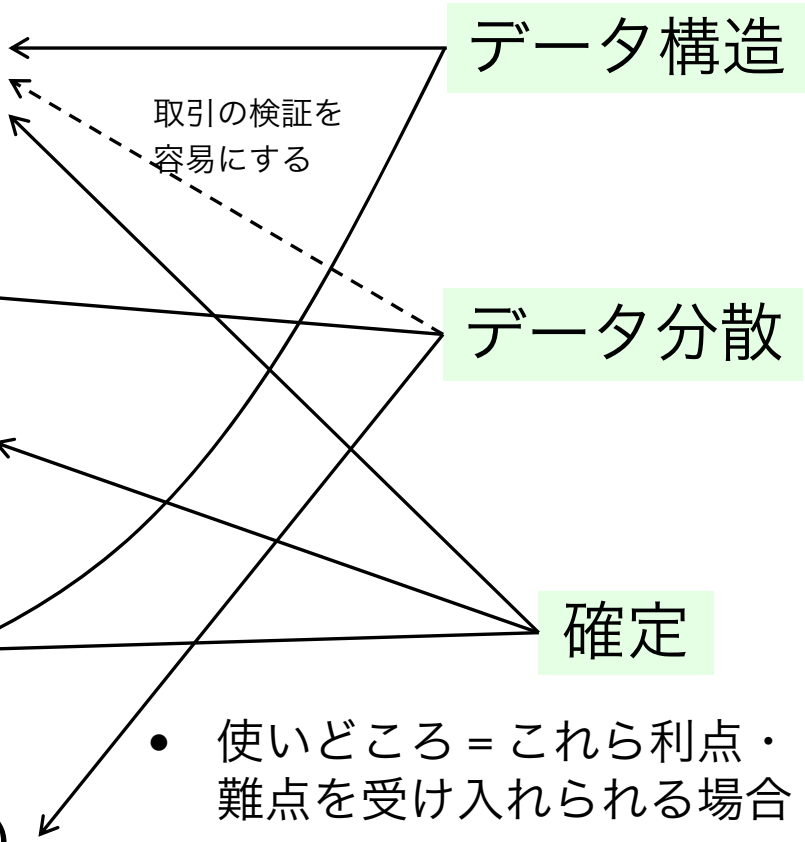
- **利点**
  - 改ざん困難・不整合 (二重消費) 防止
  - 耐故障性
  - ノード数不明＆多数でも確定可
- **難点**
  - 確定に割と時間がかかる
  - (追記しかできない)
  - (性能も容量もスケールアウトしない)



利点と難点、その由来

(Bitcoin) ブロックチェーンの **設計**

- **利点**
  - 改ざん困難・不整合 (二重消費) 防止
  - 耐故障性
  - ノード数不明＆多数でも確定可
- **難点**
  - 確定に割と時間がかかる
  - (追記しかできない)
  - (性能も容量もスケールアウトしない)

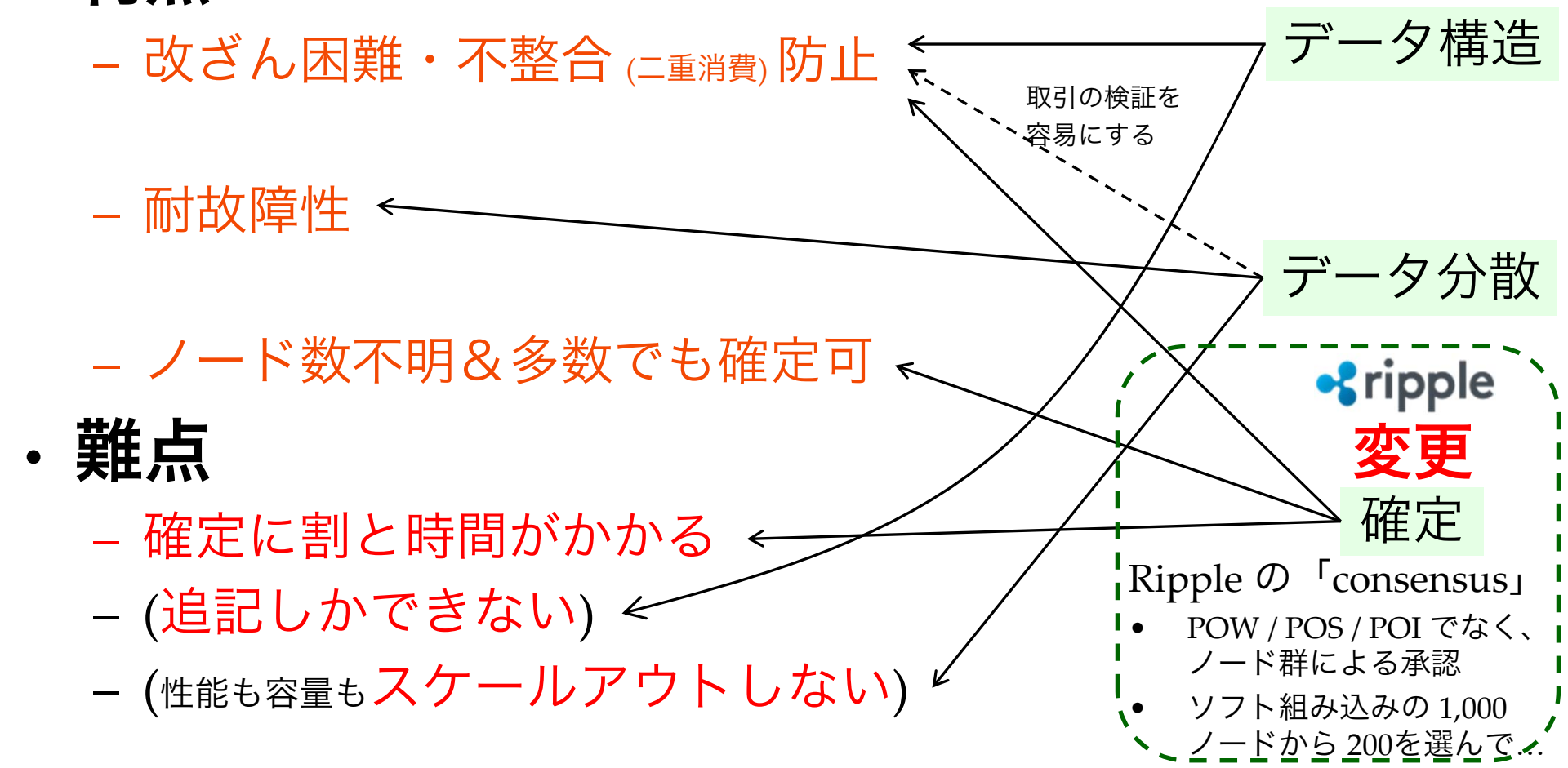


- 使いどころ＝これら利点・難点を受け入れられる場合
- 難点解決の試み (次ページ以降)

他の設計の可能性

(Bitcoin) ブロックチェーンの **設計**

- **利点**
  - 改ざん困難・不整合 (二重消費) 防止
  - 耐故障性
  - ノード数不明＆多数でも確定可
- **難点**
  - 確定に割と時間がかかる
  - (追記しかできない)
  - (性能も容量もスケールアウトしない)



他の設計の可能性

- 確定までの時間をもっと短くしたい
  - Bitcoin 10分、NEM 1分、Etheream 15秒、Ripple 5〜10秒、…
  - 普通のデータベースでは一瞬　〜 1秒

- **確定** の方法：分散合意プロトコル
  - Paxos (in Google's Chubby) や RAFT
    - ○ 一瞬で確定できる → 確定の頻度任意
    - △ ノード数 せいぜい 2桁までか
    - △ 皆がノード数を知っている必要がある
      - Paxos の手順："If the proposer receives the requested responses from a majority of the acceptors, then it can issue a proposal with number n and value v …"
      - ノードの増減が大変。全体を一時止める必要ありそう。
      - 停止させてのメンテナンスができるなら OK ？？？
      - プライベート型 / コンソーシアム型ブロックチェーンなら OK ？

Leslie Lamport 氏
- Lamport's logical clocks
- Byzantine 将軍問題
- Paxos
- …

他の設計の可能性

- 全ノードが全ブロックのコピーを持つのはばからしく見える。
  - Bitcoin 2桁 GB

- **データ分散**　の方法：各ノードは担当ブロックだけ持つ？
  - 方法：コンシステントハッシング？　分散ハッシュ表？
  - 取引の検証が難しくなる。
    軽量化プロトコルのアイディアを使える？

  - 研究の余地はあるかも。ないかも。

他の設計の可能性

- ブロックのハッシュチェーンとは異なる **データ構造** ？
  - 追記のみでは困る場合
  - どうしても 書き込むデータを小さくできない場合
    - Bitcoin の総データ量はせいぜい 2桁 GB, 2009年～
    - ハッシュ値やポインタ (URL 等) では済まない場合
  - もはや「ブロックチェーン」ではなくなるか？
  - …たぶん、ブロックチェーンを使うべきではない
    - 分散データベースなり何なり。

まとめ

- 利点・難点がある：
  - 例：改ざん困難・不整合防止 vs. 追記のみ
  - 使いどき ＝ これらを受け入れられる場合
- それらを裏付ける設計がある：

  データ構造　　データ分散 の方法　　確定 の方法

- 難点を解決する試みが続いている
  - 特に「確定」の方法
- おまけ：インターネットのような
  社会基盤となるためには？　特に技術面