LSTMを用いたGANによる疑似トラヒックの生成に関する一考察

長岡技術科学大学　電気電子情報工学課程4年
通信ネットワーク研究室　栗山海渡

2020年 9月26日

目的・背景

- 近年，ユーザが利用する端末が多様化
- トラヒックジェネレータを用いて，試験用の疑似トラヒックデータを生成し，シミュレーションやテストを行う

  ex）キャパシティプランニング
  - ITシステムの構築の際に使われる
  - あるトラヒックを処理するのに，どのくらい増強すれば良いか
- トラヒックジェネレータにおける問題
  - 公開されているデータセットが少ない
  - リアルなトラヒックを作るのが困難
  - 統計学的知識やパラメータ設定が必要

目的・背景

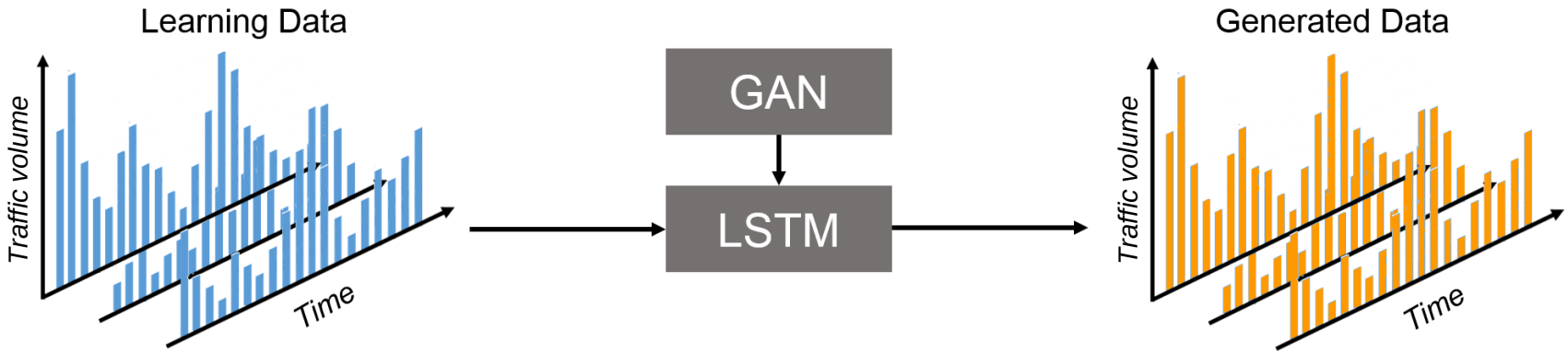
■ 目的

目指すところ

一つのフローのトラヒック特性から類似したトラヒックを生成

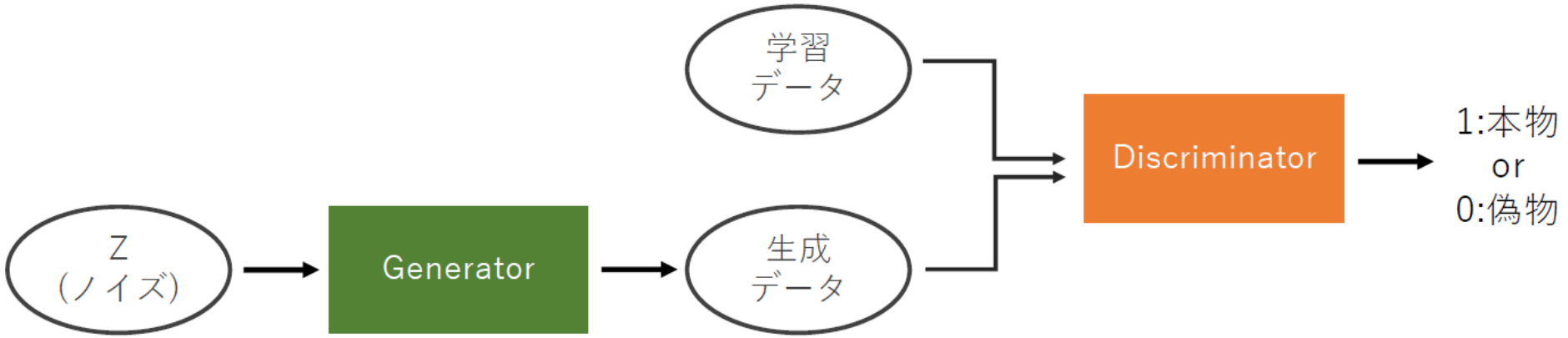
多数のフローのトラヒック特性から相関を再現したトラヒックを生成

あるネットワークのトラヒックから別のネットワークのトラヒックを生成



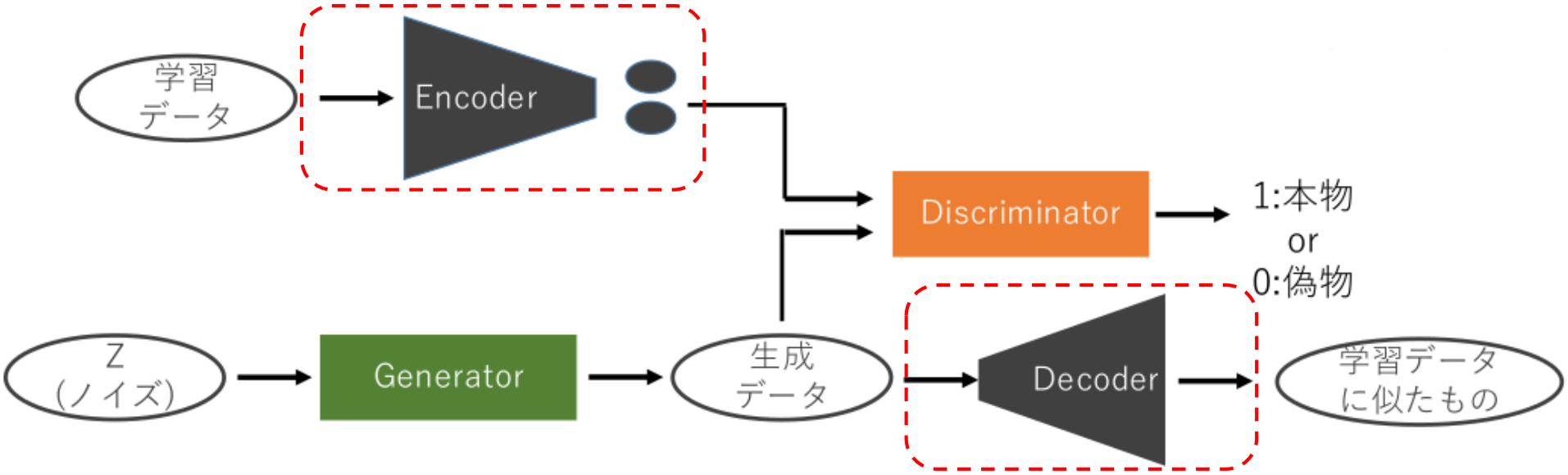
GAN（Generative Adversarial Networks）

- 本物のデータに類似したデータを生成する教師なし学習モデル
- GANの構成
  - 識別器(Discriminator)
    - 入力されたデータが，学習データか生成データかを識別
    - 学習データなら1，生成データなら0となるように学習
  - 生成器(Generator)
    - Discriminatorを騙せるほど類似したデータを生成
    - Discriminatorの出力を1に近づけるように学習



先行研究

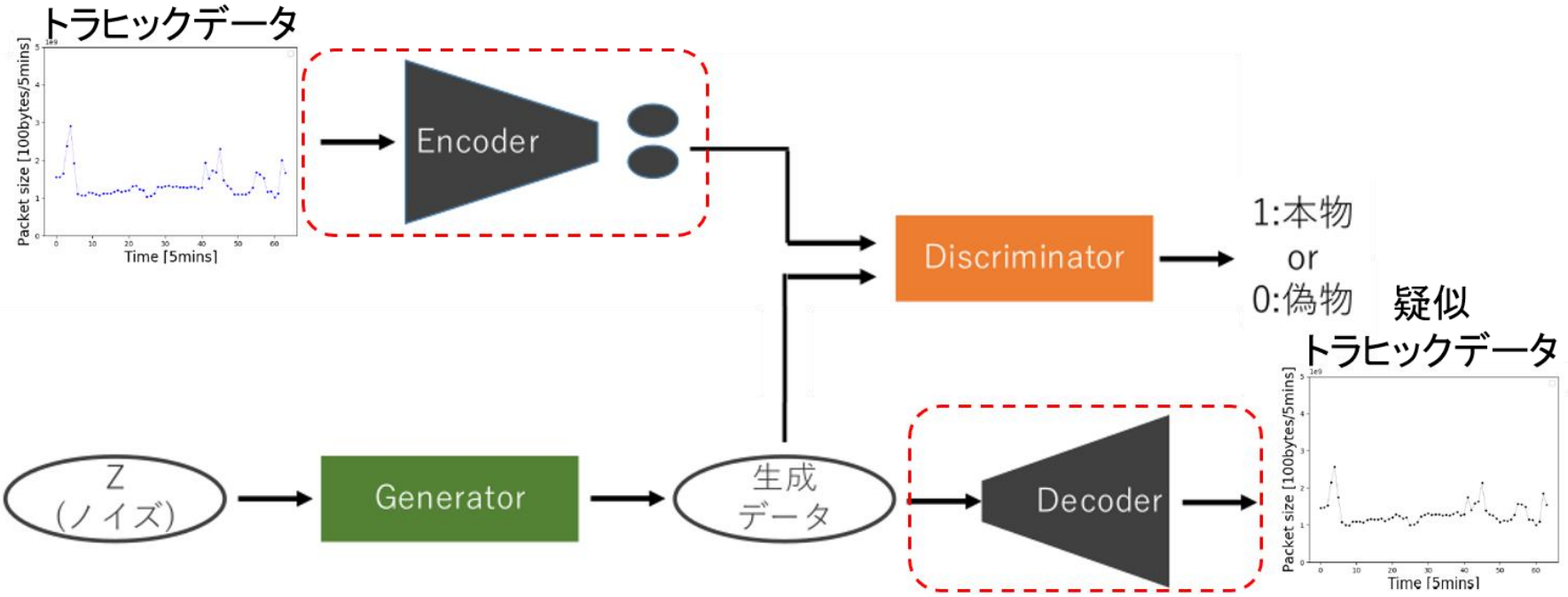
- GANとAutoEncoderを組み合わせた疑似トラヒック生成手法
- AutoEncoderの構成
  - Encoder ・・・ データの次元圧縮を行う
  - Decoder ・・・ 次元圧縮されたデータを復元する
- 次元圧縮することでGANの生成範囲を狭めることができる



➡ 学習データと生成データの長さを変更することができない

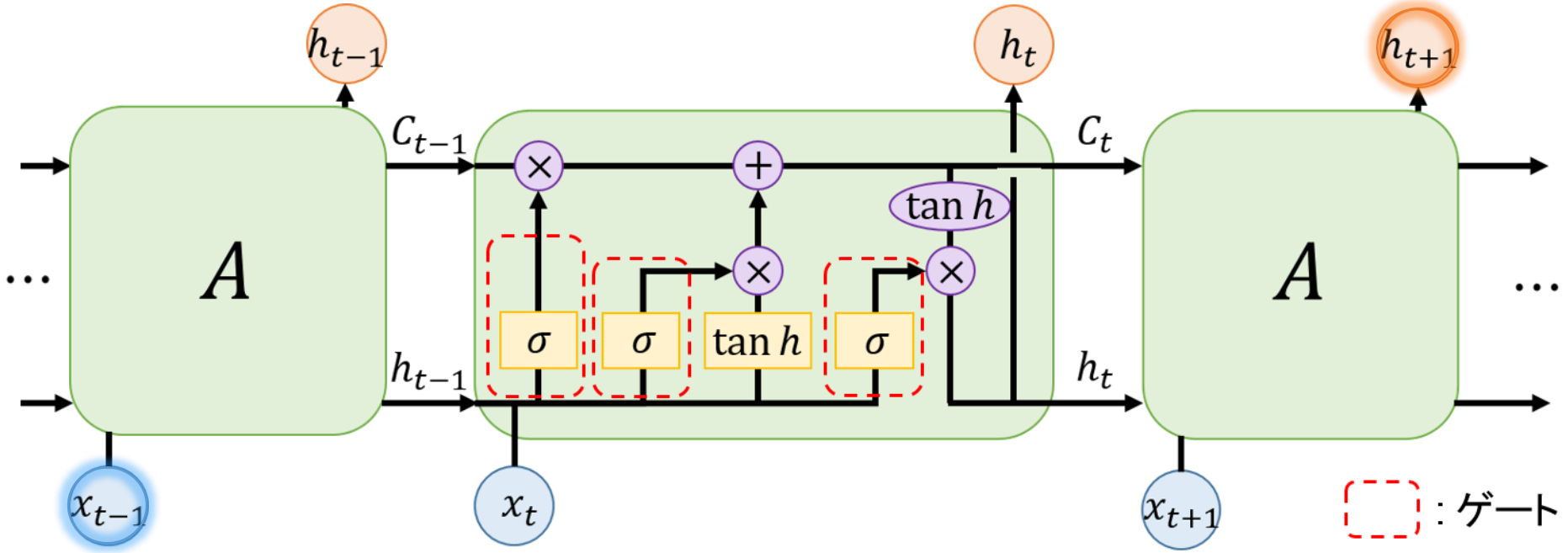
先行研究

- 学習方法
  1. トラヒックデータをAutoEncoderに学習させる
  2. 学習済みのEncoderとGANを組み合わせ，GANのみを学習させる
  3. Generatorからの生成データをDecoderに入力し，疑似トラヒックデータを生成する



LSTM(Long Short Term Memory)

- 長期的な依存関係を学習できるため，時系列データに対してよく用いられる
- LSTMの構造
  - セル(C)・・・セル状態に情報を追加または削除していく
  - ゲート・・・シグモイド層(σ)によって，0～1の数値を出力することで，前の情報(h)を引き継ぐ

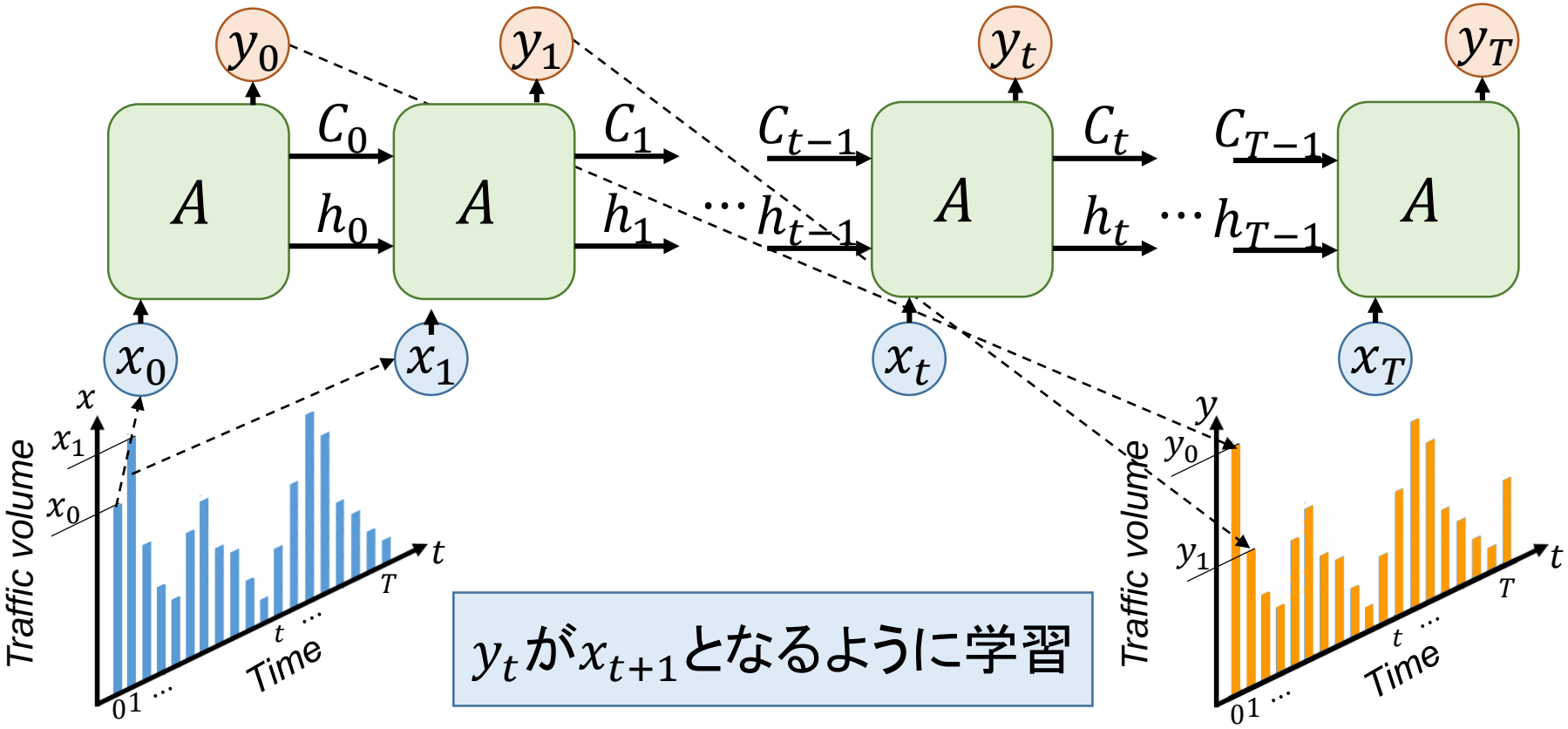


提案手法

- GANとLSTMを組み合わせた疑似トラヒック生成手法

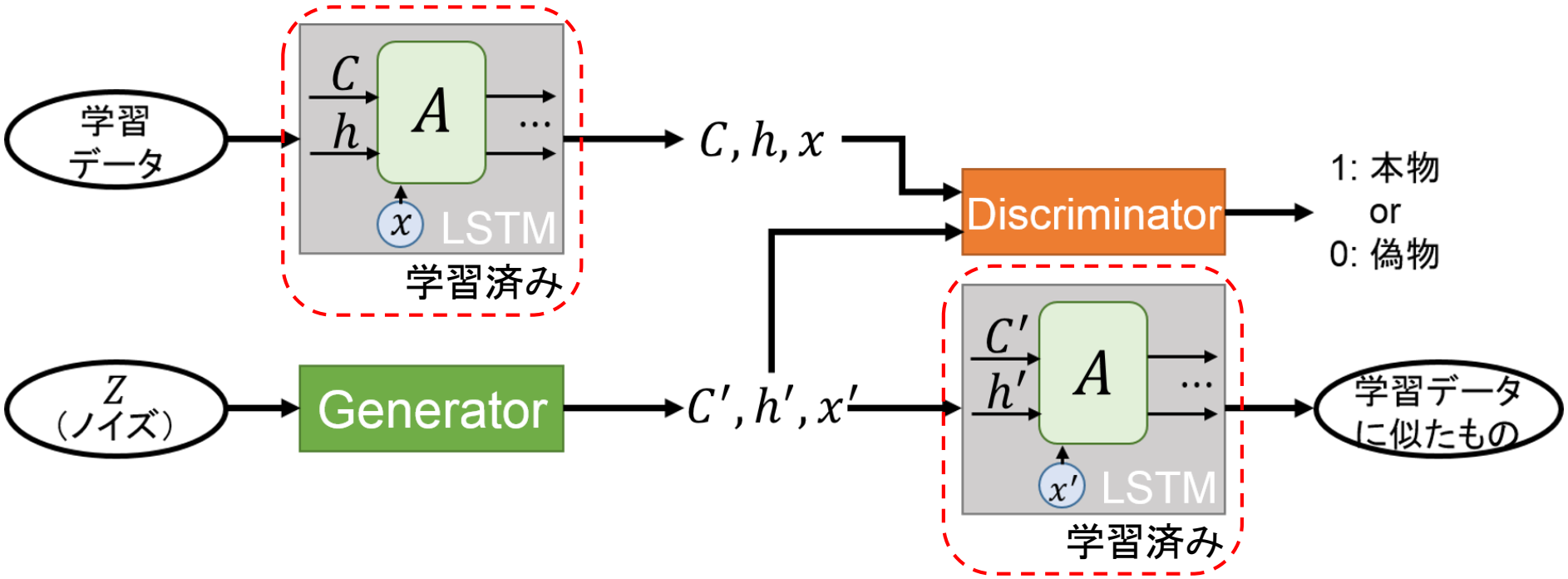
➡ 学習データと生成データの長さを任意に変更可能

  - LSTMにより，トラヒックデータの時系列的な特性を学習



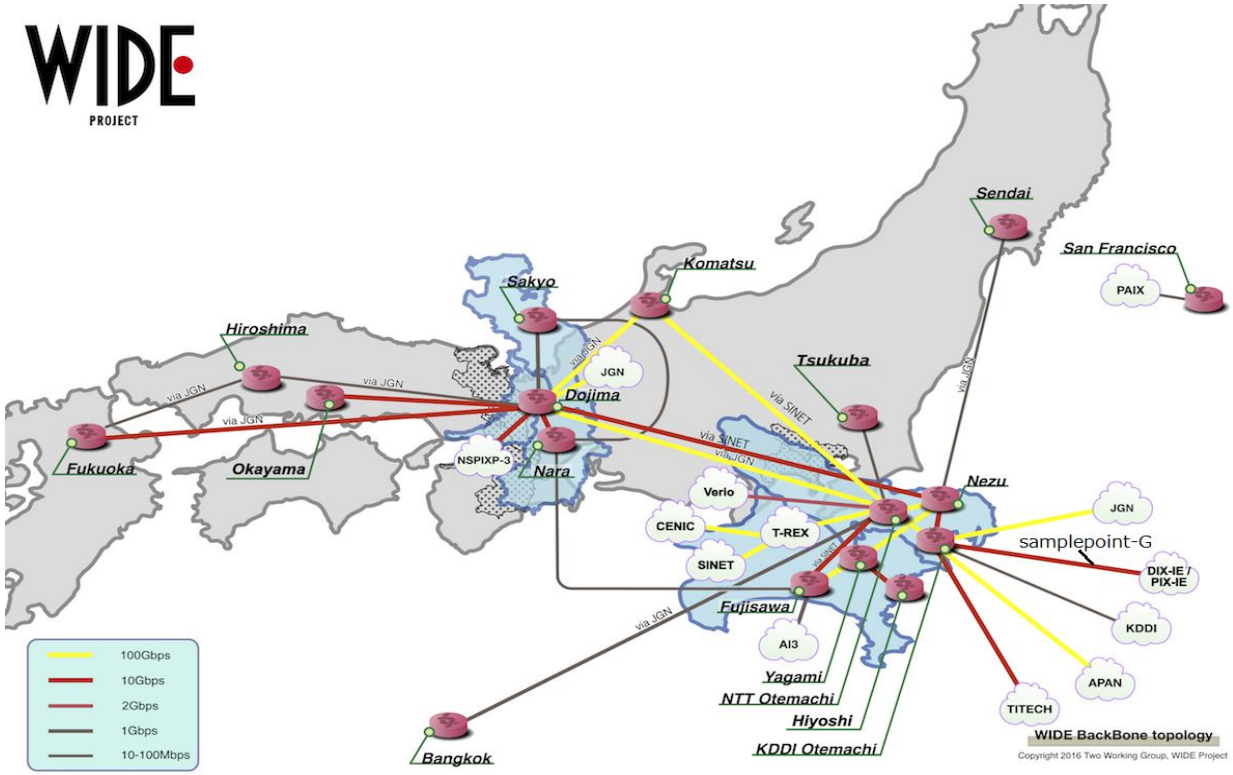
提案手法

- GANとLSTMを組み合わせた疑似トラヒック生成手法
  - GANにより，学習済みのLSTMに入力する最初のセル状態Cと隠れ層h，入力xを学習
  - 学習済みのGeneratorとLSTMにより，疑似トラヒックデータを生成



データセット

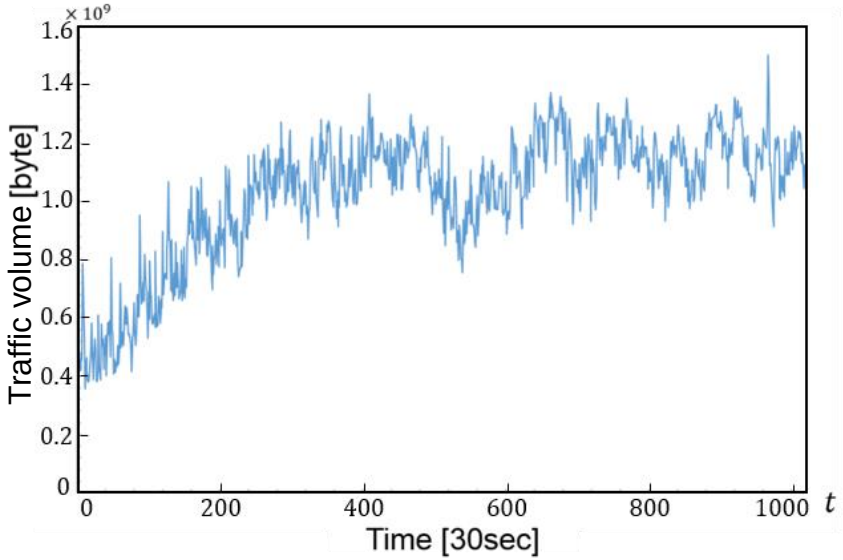
- MAWI Working GroupのWIDEネットワークトポロジ[2]



- 複数の点で測定された長期的なネットワークトラヒックを公開
- サンプリング点Gにおける2020年4月20日の8時間のトラヒックトレースを使用

シミュレーション

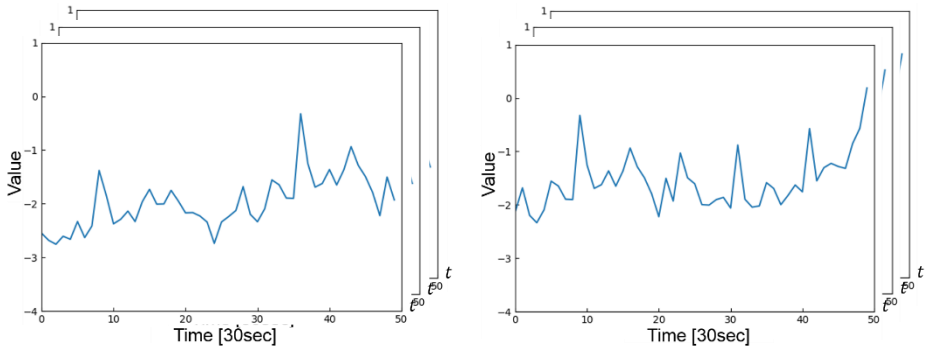
- データセット
  - ある1日の30秒毎に何バイト通信されたかを表したデータ
- 学習データ
  - 前処理として，標準化を行う（平均値＝0，標準偏差＝1）
  - tを1ずつずらして学習データを生成（$t = 0{\sim}50, 1{\sim}51, \cdots$）



データセット

訓練データ
（80個）

テストデータ
（20個）



学習データ（100個）

シミュレーション

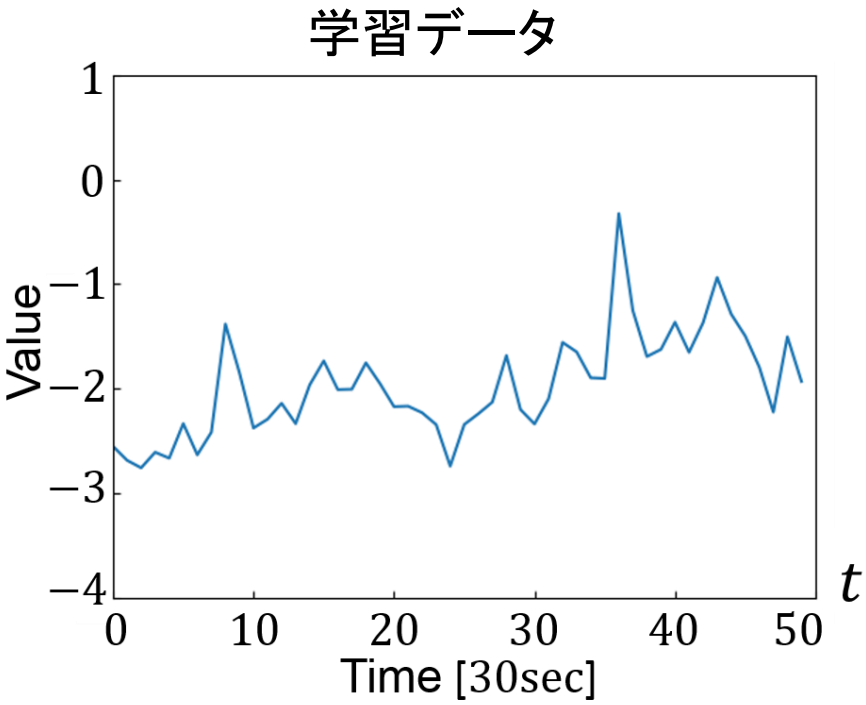
- GANのニューラルネットワークの構成

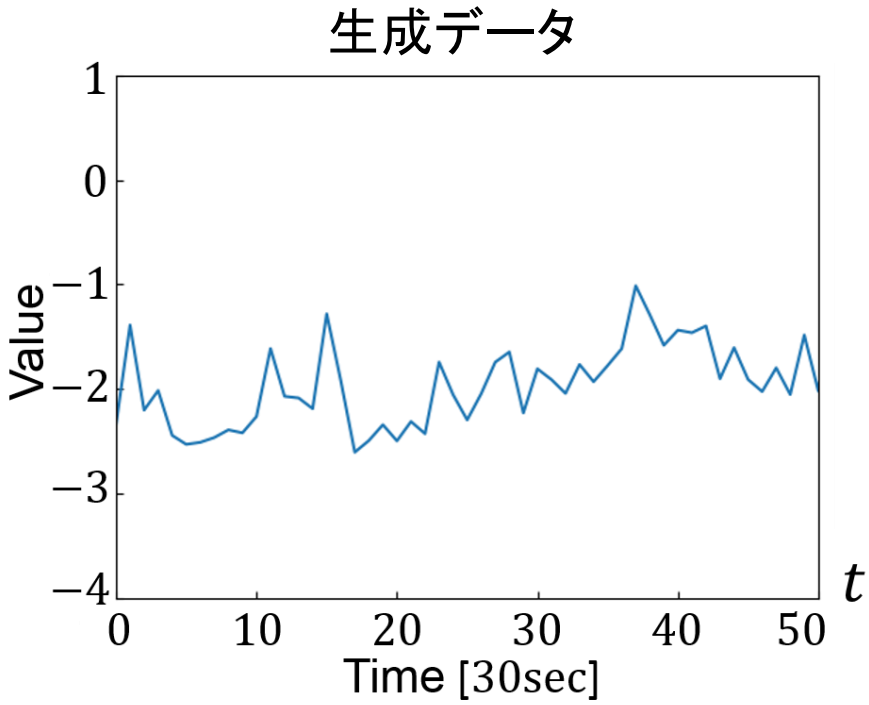
	Generator		Discriminator	
Layer	Units	Act.	Units	Act.
Input	100	-	(1, 1, 10)	-
Hidden	128	LReLU	512	LReLU
Hidden	256	LReLU	256	LReLU
Hidden	512	LReLU	-	-
Hidden	1024	LReLU	-	-
Output	(1, 1, 10)	-	1	Sigmoid

  - エポック数: 300
- LSTMの各パラメータ
  - 隠れ層の次元: 5
  - バッチサイズ: 4
  - エポック数: 100000

結果

- 生成結果の一部





- 挙動が似ていることが確認できる
- 同様の挙動をとる異なるトラヒックデータが生成できている

⇒ 元のトラヒックと類似したトラヒックデータを生成できた

まとめ・今後の予定

- まとめ
  - バイトレベルにおけるLSTMとGANを組み合わせた新たな疑似トラヒック生成手法を提案
  - 元のトラヒックと類似した疑似トラヒックデータを生成することができた

- 今後の予定
  - LSTMとGANの各パラメータの調整
  - 評価手法の検討
  - データの前処理の検討

参考文献

[1] 山際哲哉，渡部康平，中川健治，敵対的生成ネットワークを利用した疑似トラヒック生成に関する一考察．信学技報，Vol. 119, No. 125, pp. 27–29, 2019.

[2] Kenjiro Cho, Koushirou Mitsuya, and Akira Kato. Traffic data repository at the wide project, USENIX Association, pp. 51, 2000.

補足

- トラヒックジェネレータは大きく三つのレベルに分類される