



# 敵対的生成ネットワークを利用した疑似トラフィック生成に関する一考察

A Study on Pseudo Traffic Generation Using Machine Learning

山際 哲哉 渡部 康平 中川 健治

長岡技術科学大学 大学院工学研究科 電気電子情報工学専攻

## 研究背景・目的

- ◆ ネットワークを構築する際には、トラフィックジェネレータを使用し、負荷テストを行うことは重要である
- ◆ トラフィックの特性を多面的に再現した疑似トラフィックの生成は困難
- ◆ 本研究では、機械学習の敵対的生成ネットワークを用いて、疑似トラフィックデータの生成法を提案し、妥当性を検証する

## 敵対的生成ネットワーク(Generative Adversarial Networks;GAN)

- ◆ GAN[1]は学習させた画像から、類似した画像を生成する機械学習
- ◆ 識別器(Discriminator)と生成器(Generator)の二つのニューラルネットワークで構成されている

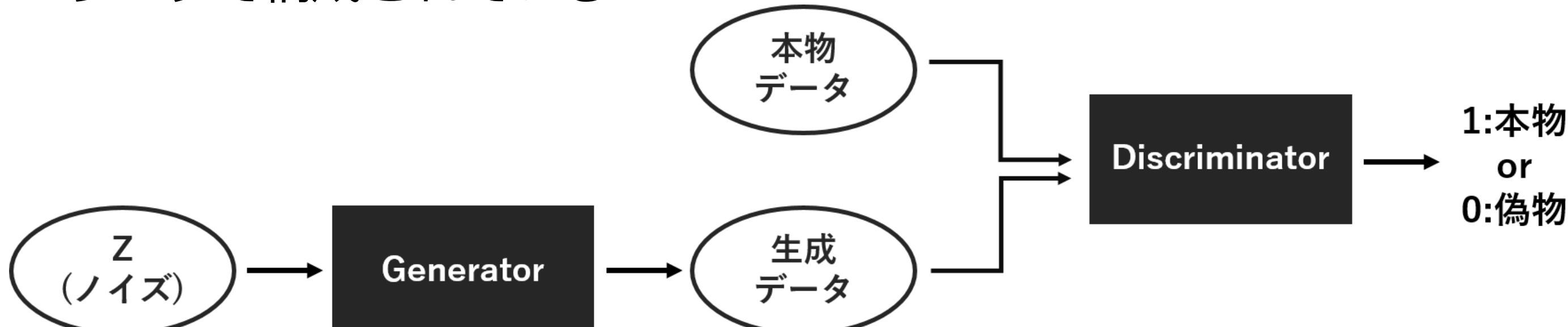


図1 GANの構成

### Discriminator

- ◆ 本物のデータとGeneratorからの生成データを識別する
    - ◆ 本物データが入力されたらを1を出力するように学習する
    - ◆ 生成データが入力されたらを0を出力するように学習する
- ⇒ Discriminatorの出力は本物のデータである確率を意味する

### Generator

- ◆ Discriminatorを騙せるほど本物のデータに類似したデータを生成する
  - ◆ 生成データをDiscriminatorに入力したときの出力が本物のデータである確率に近づけるように学習する
- ◆ DiscriminatorとGeneratorが敵対し学習し合うことで、Generatorは本物のデータに類似したデータを生成する

## 疑似トラフィック生成のためのGAN

- ◆ 本研究で扱うトラフィックデータは、計測期間について一定時間間隔毎に送信されたバイト数の時系列である
- ◆ 図2のように、トラフィックデータをニューラルネットワークに入力し、疑似トラフィックデータを出力する
  - ◆ 入力：時間間隔毎のバイト数を成分に持つベクトル
  - ◆ 出力：疑似トラフィックデータを表す同様のベクトル

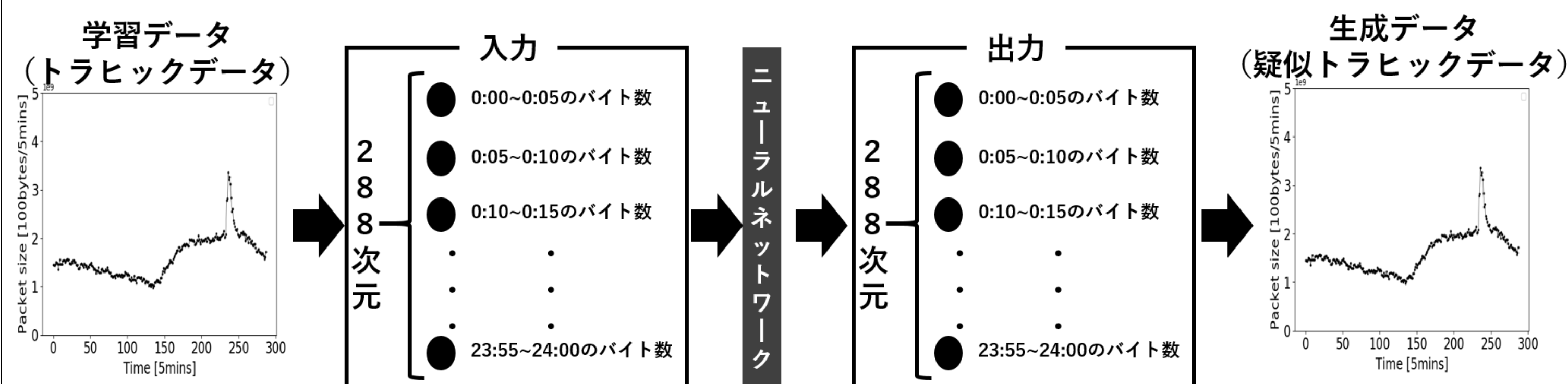


図2 トラフィックデータの入出力の例

### 提案手法

- ◆ 次元圧縮手法であるEncoderを用いる
    - ◆ Encoderによってトラフィックデータの特徴を捉える
    - ◆ Encodeされたトラフィックデータの特徴を潜在空間に落とし込む
- ⇒ GANの学習及び生成を容易にし、疑似トラフィックデータの生成が可能

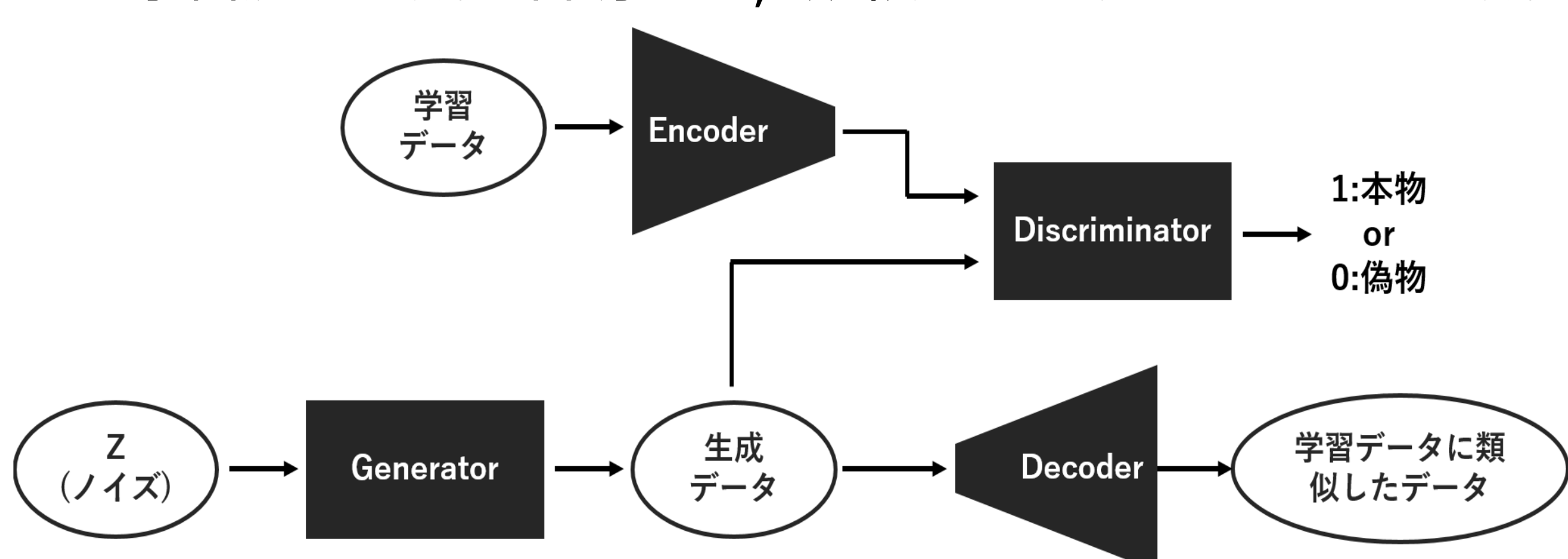


図3 疑似トラフィック生成のためのGANの構成

### 学習及び生成方法

- ① トラフィックデータをEncoder,Decoderに学習させる
- ② EncoderとGANを組み合わせて学習(Encoderは学習させない)
- ③ Generatorからの生成データをDecoderに入力し、疑似トラフィックデータを生成する

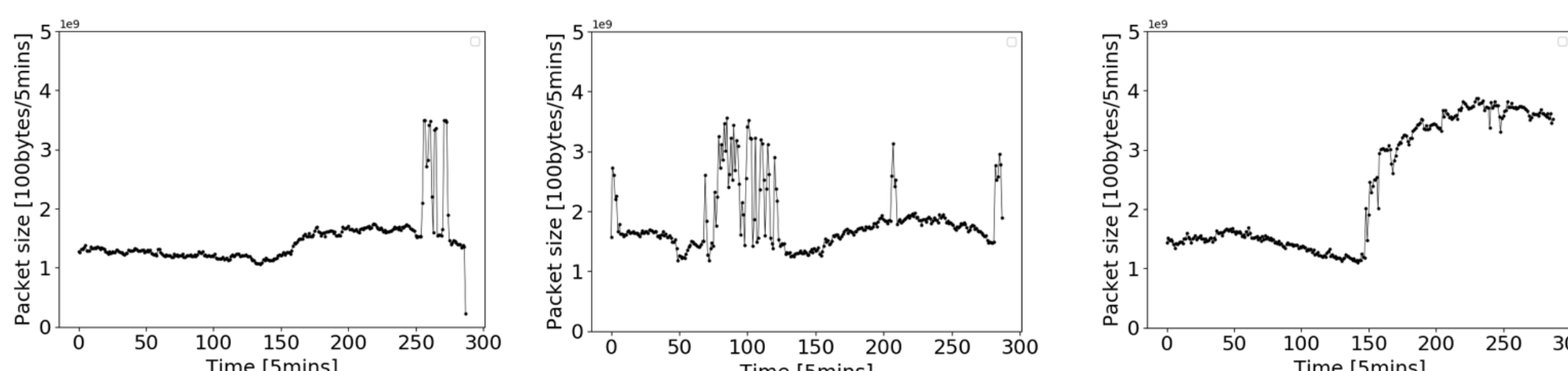
## トラフィックデータ生成実験

### 方法

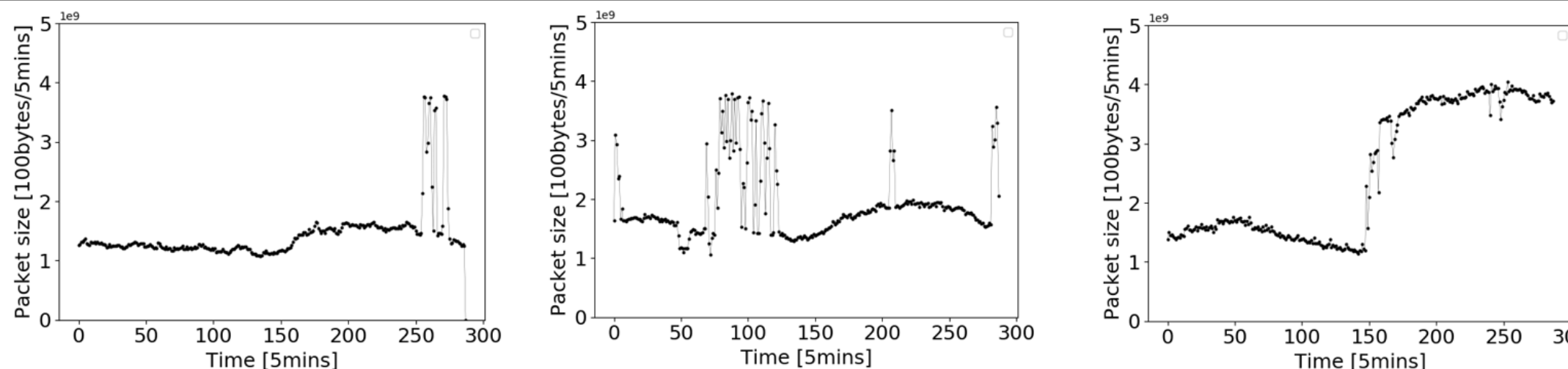
- ◆ 公開されているトラフィックデータ[2]を学習データとした
  - ◆ 5分間毎のバイト数、1日分を1つの学習データで168日分ある
- ◆ コルモゴロフ-スミルノフ検定(KS検定)で評価した
  - ◆ 二つの確率分布が一致していないかを検定する
  - ◆ 5分間毎のバイト数における学習データと疑似トラフィックデータをそれぞれ168サンプルで検定する

### 結果

生成データ  
(類似)



学習データ



生成データ  
(一部類似)

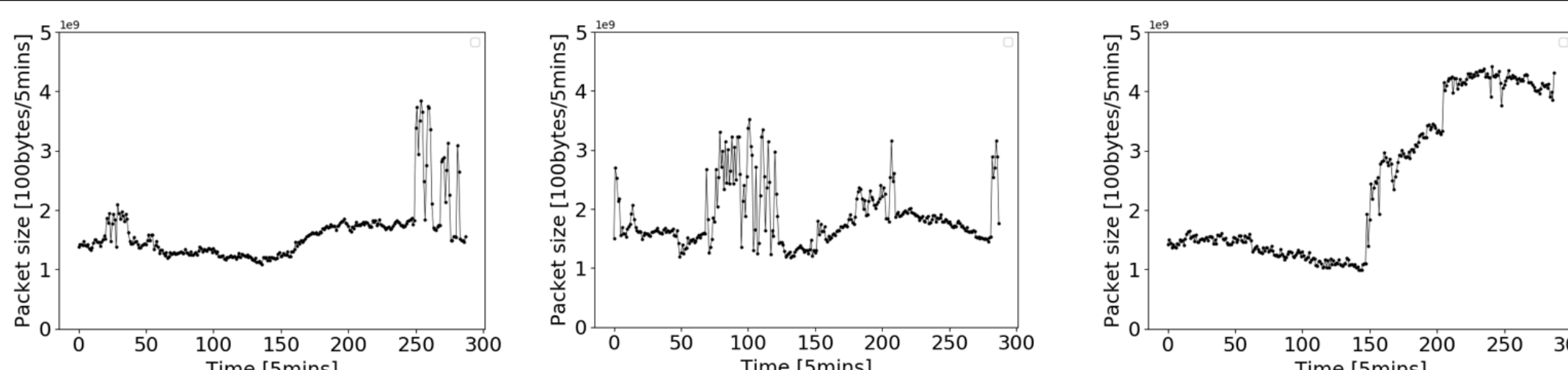


図4 学習データと生成データの例

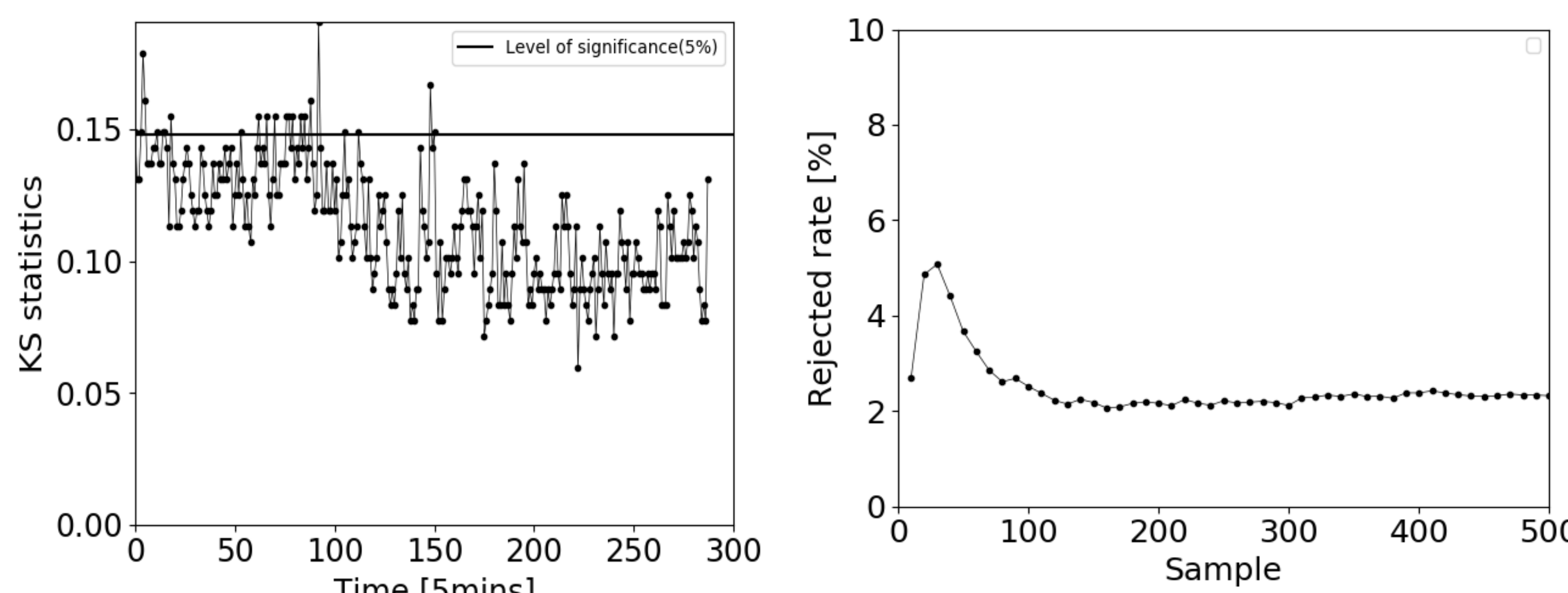


図5 KS検定の結果

図6 棄却される割合

- ◆ 学習データに類似した生成データであることが確認できる
- ◆ KS検定の結果から、多数の点が棄却されていない
- ⇒ 生成データは本物のデータの分布とかけ離れてはいない
- ◆ GANはランダムな生成のため、必ずKS検定で棄却される訳ではない
- ◆ GANで500サンプル生成し、棄却される割合を確認した結果、2.3%に収束した
- ◆ 本研究では有意水準を5%に設定しており、実データで検定を行った場合、95%棄却されないが、GANの場合97.7%棄却されていない
- ⇒ GANで生成した疑似トラフィックデータは本物のトラフィックデータとほぼ同等であることを確認した

## まとめ

- ◆ GANを用いて疑似トラフィックデータの生成法を提案し、妥当性を検証した
- ◆ 今後は、GANの精度向上及び生成されるトラフィックに多様性を持たせられるかを確認する

## 参考文献

- [1] I. Goodfellow et al., “Generative Adversarial Nets,” Advances in Neural Information Processing Systems 27, pp.2672-2680, 2014
- [2] “Abilene Topology and Traffic Dataset”, <http://www.cs.utexas.edu/~yzhang/research/AbileneTM/>