

# V2I 네트워크 환경에서 Systematic Network Coding을 이용한 데이터 기반 고신뢰 전송 방법

권정민\*, 박형곤<sup>o</sup>

## Data Driven Reliable Dissemination Strategy Based Systematic Network Coding in V2I Networks

Jungmin Kwon\*, Hyunggon Park<sup>o</sup>

요약

본 논문에서는 RSU(Road Side Unit)와 OBU(On-Board Unit)로 구성된 V2I(Vehicle-to-Infrastructure) 환경에서 교통 데이터를 이용하여 SNC(systematic network coding) 기반 고신뢰도 데이터 전송 시스템을 제안한다. RSU에서 빠르게 이동하는 OBU로 데이터를 전송할 경우 RSU에서 데이터를 전달할 수 있는 시간은 OBU의 속도에 의하여 결정되기 때문에, 데이터의 최대 정보량은 시간에 따라 변하는 차량의 속도에 결정되는 특징이 있다. 따라서, 본 논문에서는 OBU의 속도에 맞추어 최대의 정보량으로 전송하면서 데이터 디코딩 에러율을 감소시키기 위하여 회귀 분석 모델을 이용한 SNC 시스템을 제안하였고 실제로 수집된 교통 데이터를 이용하여 성능을 확인하였다. 다양한 회귀 분석 알고리즘을 이용한 제안한 SNC 시스템은 기존 방법과 비교하여 인코딩 비율은 유지하면서 디코딩 성능이 향상됨을 실험을 통하여 확인하였다.

**키워드** : 시스티메틱 네트워크 코딩, 차량과 인프라 통신, 교통 예측, 회귀 분석 모델, 브로드캐스트

**Key Words** : SNC (Systematic Network Coding), vehicle to infrastructure (V2I) network, traffic estimation, regression model, broadcast

### ABSTRACT

In this paper, we propose a reliable data dissemination strategy based on traffic data using SNC (systematic network coding) in V2I (Vehicular-to-Infrastructure) networks that consist of RSUs (Road side Unit) and OBUs (On-Board Unit). Since the time duration of OBUs serviced by an RSU is determined by the speed of fast-moving OBUs, the maximum amount of information transmitted from RSU to OBU correspondingly changes over time. Therefore, in this paper, we propose an SNC system using regression analysis models in order to disseminate the maximum amount of information while reducing the decoding error rate, and we confirm the performance by using actually collected traffic data. Experiment results show that the proposed SNC system using various regression analysis algorithms improves decoding performance while maintaining an encoding rate compared to conventional approaches.

※ 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구이며(No. 2019-0-00024, 네트워크 자동화를 위한 개방형 네트워크 데이터 분석 기반 지도형 애자일 머신러닝 기술 개발), 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단(No. NRF-2017R1A2B4005041)과 2017학년도 이화여자대학교 대학원 장학금 지원에 의한 논문임.

♦ First Author : Ewha Womans University Department of Electronic and Electrical Engineering, jungmin.kwon@ewhain.net, 학생회원

o Corresponding Author : Ewha Womans University Department of Electronic and Electrical Engineering, hyunggon.park@ewha.ac.kr, 종신 회원

논문번호 : 201909-207-B-RN, Received September 26, 2019; Revised November 23, 2019; Accepted November 29, 2019

## I 서 론

최근 원활하고 안전한 교통이 이루어질 수 있도록 차량 또는 도로 등 교통 체계를 구성하는 요소들에 대한 실시간 정보를 전달하는 연구가 활발히 이루어지고 있다. 특히, 차량 통신은 도로 안전과 인포테인먼트(infotainment) 등 다양하고 빠른 서비스를 제공하는 기술에 초점이 맞춰져 더욱 큰 관심을 받고 있다<sup>[1]</sup>. 이에 따라 C-ITS(Cooperative Intelligent Transport System) 기술 개발이 활발히 이루어지고 있는데, C-ITS는 급변하는 교통 상황과 다양한 정보를 주행 중인 차량에 실시간으로 제공하는 기술로, 차량 간 저지연 및 고신뢰 전송 방법에 관한 연구가 매우 필수적이다<sup>[2]</sup>.

도로 위를 주행 중인 차량에 정보를 제공하는 차량 통신(Vehicle-to-Everything, V2X)은 대표적으로 차량과 차량 간의 통신(Vehicle-to-Vehicle, V2V), 그리고 차량과 인프라 간의 통신(Vehicle-to-Infrastructure, V2I)이 있으며, 그림 1과 같이 서버(server)와 백본망(backbone network), access gateway, RSU(Road Side Unit), 그리고 OBU(On-Board Unit)로 구성되어 있다. RSU와 OBU 간의 통신을 나타내는 V2I 통신은 사용자 편의 서비스를 제공하기 위하여 다양한 정보 제공을 목적으로 하며, 높은 전송량과 서비스 시간을 요구하는 서비스들을 주로 제공한다<sup>[3]</sup>. 이러한 V2I 시스템을 위한 대표적 통신 기술로는 DSRC(Dedicated Short Range Communication) 프로토콜이 있으며, 국내뿐만 아니라 세계적으로 일부 거점 도로와 자율주행 실증 도시에 DSRC 인프라를 구축함으로써 C-ITS 실현을 이끌고 있다.

V2I 네트워크는 차량의 위치와 속도 정보, 주변 교통 상황 정보를 주고받기 위하여 데이터를 전달하는데, 빠르게 변화하는 차량의 속도와 RSU 링크에 대한 짧은 연결 시간으로 인하여 패킷 손실률이 높으며 다중 접속 스케줄링과 채널 할당 문제가 발생하게 된다<sup>[4]</sup>. 이러한 V2I 환경의 무작위적인 접속과 예측 불가능한 지연 및 손실이 발생함에 따라, TDMA(Time-Devision Multiple-Access)를 이용한 MAC 프로토콜이 IEEE 802.11에서 제시되었다<sup>[4]</sup>. 동적으로 전송 시간 슬롯을 개별 차량에 할당하는 R-ALOHA(Reservation ALOHA) MAC 프로토콜이 IEEE 802.11 V2I 통신에서 제안 및 연구되어왔으나<sup>[5]</sup>, 이러한 MAC 프로토콜 방식은 차량의 GPS를 통하여 확인할 수 있는 시간과 정확한 동기화가 요구된다. 이처럼 TDMA 기반의 MAC 프로토콜은 현존하

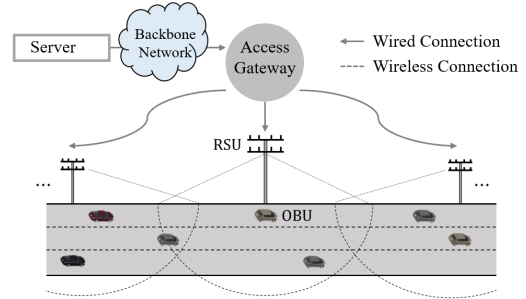


그림 1. 차량 네트워크 구조  
Fig. 1. Vehicular network structure.

는 DSRC 기술에서의 적용이 적절하지 않아 IEEE 802.11p/DSRC 중 EDCA(Enhanced Distributed Channel Access) 브로드캐스트 모드를 이용한 데이터 전달이 제안되었다<sup>[6]</sup>. EDCA 기술은 request-to-send, clear-to-send, acknowledgement(ACKs) 등의 통신 동작을 생략하기 때문에 오버헤드와 전송횟수를 줄여 데이터 전달이 완료되기 때문에 빠르게 변화하는 차량 환경에 안정적으로 데이터를 전달 가능한 장점이 있다. 하지만, V2I 네트워크 채널 상태로 인해 발생하는 손실은 여전히 고려되지 못한 한계점이 있다. 따라서 본 논문에서는 채널 손실에 따라 요구되는 RSU와 OBU 사이의 데이터 전송 효율성과 신뢰성을 모두 향상시키는 알고리즘을 제안하고자 한다.

본 논문은 NC(Network Coding) 기법 중 SNC(Systematic Network Coding)<sup>[7]</sup>를 이용하여 V2I 환경에서 교통 데이터 기반 신뢰도 높은 데이터 전달 방법에 대하여 제안한다. 차량이 빠르게 이동하고 있는 도로의 통신환경은 실시간으로 데이터를 전달하는데 어려움이 있어 재전송과 같은 추가적인 지연(delay)을 줄이는 것이 매우 필수적으로 요구된다. 이에 따라 VANET(Vehicular Ad hoc Network) 환경에서 NC의 인코딩 기법으로 가장 널리 사용되는 RLNC(Random Linear Network Coding)를 이용하여 재전송을 줄이는 데이터 전달 방법이 제안되었다<sup>[8]</sup>. 그러나 RLNC의 경우 디코딩에서 요구되는 복잡도가 매우 높은 문제가 있어 처리 지연이 발생하고 전력 소모가 높아지는 단점이 있다. 이에 따라 NC의 인코딩 기법 중 XOR(Exclusive OR,  $\oplus$ )을 이용하여 특정 데이터 그룹의 수만큼 인코딩하여 전달하는 연구가 제안되었다<sup>[9]</sup>. 변화하는 네트워크 구조에서 영상 데이터의 그룹 간의 관계에 따라 데이터를 XOR 하는 기법은 빠른 데이터 인코딩 및 디코딩을 수행할 수 있으나 데이터를 손실하였을 경우 모든 데이터를 잃어버리는

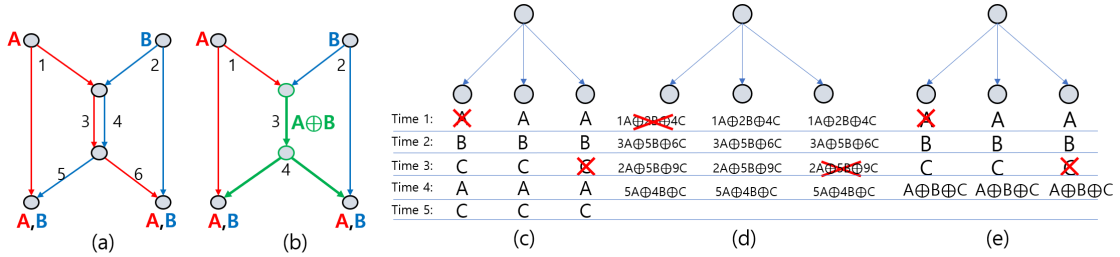


그림 2. 다양한 형태의 네트워크에서 데이터 전송 횟수 비교 (a) 나비 형태에서 기존 ‘저장 및 전달’ 방법 (b) 나비 형태에서 NC 방법 (c) 싱글 홉 네트워크에서 브로드캐스팅 전송 방법 (d) NC만을 이용하여 전송한 방법 (e) SNC를 이용하여 전송한 방법  
 Fig. 2. The number of transmission comparison in various networks (a) conventional ‘store and forwarding’ approach in butterfly network (b) NC approach in butterfly network (c) broadcasting approach in single-hop network (d) pure NC approach in single-hop network (e) SNC approach in single-hop network.

문제점이 있다. 이에 따라 본 논문에서는 하나의 RSU에서 다수의 OBU에 전달된 데이터 중 손실된 데이터에 대하여 재전송 횟수를 줄이면서 동시에 높은 데이터 복원율을 보이게 하겠다. SNC는 손실된 데이터에 대한 재전송과 디코딩 복잡도를 효율적으로 줄이며 데이터 복호 성능을 향상시킴으로써 실시간 전송 환경에 매우 적합한 통신 기법으로 알려져 있다<sup>7)</sup>. 이에 따라 V2I 환경에서 패킷 손실을 고려한 SNC 알고리즘이 제안되었으나<sup>10-11)</sup>, 차량의 환경을 잘 반영하지 못하여 차량의 속도 변화에 따른 전달 가능한 정보량을 고려하지 못한 한계점이 있다.

한편, V2I 네트워크에서 제공되는 차량의 정보는 차량의 속도, 수, 도로 구간, 도로의 종류 등 차량 환경에 한정되어 있고 통신환경이 반영된 데이터는 수집하기 어렵거나 사용하기 어려운 경우가 많은 문제점이 있다. 이에 따라서 현재 제공된 차량의 환경 데이터를 통해 RSU와 다수의 OBU 간의 예측되는 전달 가능한 정보량을 계산한 후에 해당 정보량만큼 데이터를 전달하는 과정이 필요하다. 이에 따라 본 논문에서는 실제 도로 환경에서 공개된 차량 데이터 중, 속도 데이터를 이용하여 전달 가능한 정보량 데이터를 간접적으로 생성한 후 시간에 따라 변하는 속도를 통하여 정보량을 예측한다. 이후, 예측된 정보량에 따라 SNC의 인코딩 과정에서 함께 인코딩되는 데이터의 개수를 결정된 후 전송하는 방법에 대하여 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 NC(network coding)에 대한 기본 이론에 관하여 서술하고 3장에서는 제안하고자 하는 시스템 모델과 세부 사항에 관하여 서술한다. 4장에서는 본 연구의 문제와 해결 방법을 정의하고 5장에서는 해당 시스템의 성능을 확인하고자 실험 설계 내용을 서술하고 결과

에 따른 시스템 성능을 분석하며, 마지막으로 6장에서 결론을 맺는다.

## II. Network Coding

### 2.1 Network Coding의 개념

NC는 기존 ‘저장 및 전달(store and forwarding)’ 방식의 전통적 네트워크 데이터 전달 방법과 달리, 각 노드에 입력되는 데이터 간의 수학적 연산을 통하여 연산 결과 값을 전달하는 방법으로 네트워크의 전송량을 크게 향상시킬 수 있는 장점이 있다<sup>12)</sup>. 이러한 NC의 장점은 나비 형태 네트워크(Butterfly Network)에서 두드러지게 나타난다. 그림 2-(a)와 같이 두 개의 발신 노드에서 중간 노드를 통하여 데이터 A, B를 수신 노드로 전달할 경우, 중간 노드 사이의 병목 현상으로 인하여 전송이 지연되거나 전송횟수가 증가하게 되어 네트워크의 효율이 감소하는 문제가 발생한다. 그러나 중간 노드에서 A와 B 사이의 XOR 연산을 통한 NC 데이터를 전송할 경우, 그림 2-(b)와 같이 한번의 전송으로 수신 노드에서 원하는 데이터를 전송할 수 있어 병목 현상을 해소할 수 있다. 즉, 기존의 라우팅으로는 6번의 전송으로 모든 데이터를 전송하는 데 반하여, NC를 사용할 경우 4번의 전송을 통하여 데이터 전달을 완료함으로써 네트워크 전송량을 향상시킬 수 있는 장점이 있다.

NC의 장점은 나비 형태 네트워크와 같은 멀티 홉 네트워크뿐만 아니라 다량의 수신 노드가 존재하는 싱글 홉(single-hop) 네트워크에서도 나타난다. 싱글 홉 네트워크는 그림 2-(c)와 같이 하나의 발신 노드에서 다량의 수신 노드가 싱글 홉으로 연결되어있고 모든 수신 노드에 서로 다른 데이터를 모두 전달하고자 한다. 브로드캐스트(broadcast)를 이용한 데이터 전달

은 유니캐스트(unicast) 또는 멀티캐스트(multicast)와 달리 수신 노드의 IP/MAC 주소를 공유하지 않고 동일 정보를 한꺼번에 전송할 수 있다는 장점이 있다. 그러나 데이터를 브로드캐스트로 전달하였을 때 특정 패킷을 수신 노드에서 잃어버렸을 경우 네트워크 전송량이 매우 감소하는 문제가 발생하게 된다. 따라서, 이러한 문제점을 해결하기 위하여 아래와 같이 크게 3 가지 방법으로 데이터를 복구한다.

**Broadcasting<sup>[13]</sup>:** 그림 2-(c)와 같이 A, B, C의 데이터를 차례로 전송하게 되고, 이때 손실된 패킷에 대해서는 ARQ(Automatic Repeat Request)를 통하여 재전송한다. 따라서 총 5번의 전송이 필요하게 된다.

**Pure Network Coding<sup>[12]</sup>:** 그림 2-(d)와 같이 모든 데이터를 NC를 이용하여 인코딩한 후 전송한다. 모든 노드에서 서로 다른 패킷을 잃어버리더라도 이후에 전달받은 패킷을 이용하여 데이터를 복구한다. 그 결과, 총 4번의 전송으로 모든 데이터를 전달받는다. 그러나 인코딩 및 디코딩 과정에서 요구하는 복잡도가 높은 단점이 있다.

**Systematic Network Coding<sup>[7]</sup>:** 그림 2-(e)와 같이 기존 브로드캐스팅과 NC를 혼합하여 전송하는 방법이다. 먼저 원본 데이터를 전송한 후, 잃어버린 패킷에 대한 복구로 NC된 데이터를 추가로 전송한다. 따라서 총 4번의 전송으로 모든 데이터를 전달받으며 비교적 간단한 연산을 통하여 데이터를 복구한다.

즉, 나비 형태 네트워크와 마찬가지로 싱글 홉 네트워크에서 NC를 이용하여 데이터를 전달할 경우 전송 횟수를 낮춤으로써 네트워크 전송량을 높일 수 있게 된다. 그러나, 앞의 예시에서도 확인할 수 있는 것과 같이, NC만을 사용하여 데이터를 전달하는 pure network coding의 경우에는 모든 데이터에 대하여 인코딩 및 디코딩을 수행함으로써 높은 연산의 복잡도를 요구하게 되는 단점이 있다. 이를 해결하기 위하여 원본 데이터를 전송한 후 NC된 데이터를 추가로 전달하는 SNC가 제안되었으며, 그 결과 연산의 복잡도를 낮추며 동시에 전송 효율을 높일 수 있게 되었다. 따라서 싱글 홉 네트워크에서 브로드캐스팅을 통하여 데이터를 모든 노드에 공유하고자 할 경우, NC를 이용한 SNC 방식이 널리 이용되게 되었으며, 다음 장에서는 SNC의 인코딩 및 디코딩 기법에 대해서 자세히 설명하도록 한다.

2.2 Systematic Network Coding

하나의 소스 노드와 다수의 클라이언트 노드로 구성된 싱글 홉 네트워크를 가정하자. 원본 데이터는

NC를 수행할 수 있는 소스 노드에서 다수의 클라이언트로 모두 공유된다. 원본 데이터는 코딩 계수(coding coefficients)와 갈루아 필드(Galois Field, GF)에서 선형으로 결합된다. 필드의 크기가  $2^M$ 인 갈루아 필드를  $GF(2^M)$ 이라고 표현하며, NC와 관련된 모든 연산은 갈루아 필드에서 정의된 연산을 사용한다. 원본 데이터의 집합  $S$ 에 대하여 소스 노드는  $L$ 개의 원본 데이터 원소를 가지는 데이터 블록(block)  $X_n (\subset S) = [x_{n1}, \dots, x_{nL}]$ 으로 구분시킨다. 여기서  $x_{nl}$ 은  $n$ 번째 블록의  $l$ 번째 데이터를 나타낸다. 본 장에서는 SNC의 기본이 되는 NC의 인코딩 과정을 설명한 후, 본 논문에서 사용된 SNC의 자세한 연산 방법에 대하여 설명하고자 한다.

NC의 인코딩 과정에서는  $K$ 개의 블록들이 함께 인코딩된다. 여기서  $K$ 를 인코딩 넘버(encoding number)라고 명명하고  $K$ 개의 데이터 블록이 하나의 NC 집합을 이룬다<sup>[14]</sup>. RLNC는 다음과 같은 연산을 수행하며, 이는 앞서 언급한 pure network coding에 해당하는 연산이다<sup>[12]</sup>.

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{K+N} \end{bmatrix} = \begin{bmatrix} c_{11} & \dots & c_{1K} \\ \vdots & \ddots & \vdots \\ c_{(K+N)1} & \dots & c_{(K+N)K} \end{bmatrix} \otimes \begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix} \quad (1)$$

여기서  $\otimes$ 는 갈루아 필드의 행렬 곱셈 연산 기호이며,  $c_{mn}$ 은  $GF(2^M)$ 에 랜덤으로 선택된 코딩 계수로, 코딩 계수 행렬  $C$ 는  $(K+N) \times K$  크기의 행렬이다. 여기서 랜덤으로 선택된 코딩 계수를 통하여 생성된 인코딩 된 데이터  $Y_1, \dots, Y_{K+N}$ 를 포함하는 데이터 패킷을 RLNC 패킷이라 명명한다.

SNC 인코딩 과정은 RLNC 인코딩 과정과 달리 코딩 계수 행렬에 항등 행렬을 추가하여 원본 데이터를 연산하는 과정이 추가되며 다음과 같은 연산을 수행한다<sup>[7]</sup>.

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_K \\ Y_{K+1} \\ \vdots \\ Y_{K+N} \end{bmatrix} = \begin{bmatrix} I_K & \\ c_{11} & \dots & c_{1K} \\ \vdots & \ddots & \vdots \\ c_{M1} & \dots & c_{NK} \end{bmatrix} \otimes \begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix} \quad (2)$$

$I_K$ 은  $K \times K$  행렬로 대각 성분이 모두 1로 구성된 항등 행렬을 나타낸다. 수식 (2)의 결과로 인코딩된

데이터  $[Y_1, \dots, Y_K, Y_{K+1}, \dots, Y_{N+K}]^T$ 를 생성하는데, 이때  $I_N$  항등 행렬로 인하여  $Y_1, \dots, Y_K$ 는 원본 데이터  $X_1, \dots, X_K$ 과 동일한 데이터가 되며  $Y_{K+1}, \dots, Y_{N+K}$ 는 임의로 선택된 코딩 계수와 선형 결합을 통하여 RLNC에 기반한 인코딩 블록을 구성하게 된다.

인코딩된 데이터  $Y$ 는 코딩 계수와 함께 패키징되어 다수의 클라이언트 노드로 전송되는데,  $Y_1, \dots, Y_K$ 를 포함하는 데이터 패킷을 systematic 패킷으로 명명한다. 따라서 SNC 인코딩 과정을 통하여 생성된 패킷은 systematic 패킷과 RLNC 패킷으로 구성된다.

한편, 데이터를 전송하는 채널은  $p$ 의 확률로 손실이 발생하는 채널이라 가정한다. 즉,  $N+K$ 의 데이터를 전달하였을 때  $p(N+K)$ 의 데이터가 클라이언트 노드에 도달하지 못할 수 있다. 따라서 클라이언트 노드에서는  $\hat{N} + \hat{K} = (1-p)(N+K)$ 개의 데이터에 대하여 NC 디코딩 과정을 수행한다.

수신된 데이터  $\tilde{Y}_1, \dots, \tilde{Y}_{\hat{N} + \hat{K}}$ 에 대하여 클라이언트 노드에서는 가우시안 소거법(Gaussian elimination)을 이용하여 디코딩하게 되고 이는 아래와 같이 인코딩 과정의 역 연산을 수행한다.

$$\begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_K \end{bmatrix} = \begin{bmatrix} \tilde{I}_{\hat{K} \times K} \\ c_{11} \cdots c_{1K} \\ \vdots \\ c_{\hat{M}} \cdots c_{\hat{N}K} \end{bmatrix}^{-1} \otimes \begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_{\hat{K}} \end{bmatrix} \quad (3)$$

여기서  $\tilde{I}_{\hat{K} \times K}$ 는  $\hat{K} \times K$  크기의 항등행렬로써  $I_K$ 에서  $K - \hat{K}$ 개의 행이 패킷 손실로 인하여 제외된 행렬이다.

그러나, 만약 수신된 데이터의 개수  $\hat{N} + \hat{K} < K$ 일 경우에는 코딩 계수 행렬이 full rank를 만족하지 못하게 되어 역행렬이 존재하지 않아 데이터를 복구할 수 없게 된다<sup>[10]</sup>. 따라서  $\hat{N} + \hat{K} \geq K$ 를 만족하고 코딩 계수 행렬이 상궤터(singular) 하지 않은 경우에 대해서만 데이터  $\hat{X}_1, \dots, \hat{X}_K$ 를 디코딩할 수 있다. 이때,  $K$ 개보다 많은 데이터가 도착하는 경우에는 디코딩의 복잡도를 줄이기 위하여 systematic 패킷을 우선적으로 사용하게 된다. 한편, 본 논문에서 고려하는 SNC는 브로드캐스팅 시스템을 고려하기 때문에 잃어버린 패킷에 대한 재전송은 하지 않는다.

다음 장에서는 V2I 환경에서 SNC를 이용한 데이터 전달 시스템에 대하여 제안하고자 한다.

### III. 고려하는 V2I 환경

본 논문에서는 그림 3과 같이 편도 여러 차로의 고속도로에 있는 RSU가 도로 위를 주행하고 있는 서로 다른 OBU에 데이터를 공유하는 구조를 고려한다. RSU는 그림 3과 같이  $d_h$  높이의 차량 검지기가 설치된 구간에 맞추어  $d_e$  간격으로 설치되어있다고 가정하며 차량과 RSU 사이의 각도는 최대  $\theta$ 가 된다.

다수의 OBU는 단일 RSU의 브로드캐스팅 범위 안에 속하게 되면 데이터를 전달받게 되는데, OBU가 이동을 하면서 RSU의 브로드캐스팅 범위를 벗어나게 되면 연속되어 설치되어있는 인접 RSU에 바로 연결한다. 한편, 본 논문에서는 고속도로에 존재하는 모든 RSU는 동기화되어있다고 가정한다. 즉, 다량의 OBU가 고속도로 내에서 이동하며 연결된 RSU가 변경될 때에 전달 데이터에 대한 정보는 모두 동기화되어있기 때문에 끊임없이 전달할 수 있다고 가정한다. 따라서 본 논문에서는 하나의 RSU와 다수의 OBU가 존재하는 하나의 단계에 대해서 고려하여 시스템을 제안하고, 이후 OBU가 고속도로를 따라 계속해서 이동할 때 발생하는 경우에 관하여 확인하고자 한다.

하나의 RSU와 다수의 OBU가 존재하는 고속도로는 시간에 따라 지속적으로 평균 속도가 변하는 특징이 있다. 이렇게 변화하는 속도는 RSU가 OBU에 전달할 수 있는 정보량의 크기를 변화시킬 수 있다. 따라서 시간에 따라 예측되는 정보량을 OBU에 적절히 전달해야, 이에 맞추어 적절히 데이터 전송을 할 수 있다. V2I 환경에서 AWGN 채널을 통하여 데이터를 전송할 수 있는 정보량을  $D$ 라고 명명한다. 이에 따라 본 환경에서 계산되는 최대 정보량  $D_T$ 는 다음과 같다<sup>[8]</sup>.

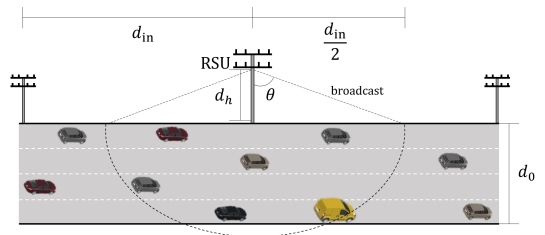


그림 3. 제안하는 V2I 네트워크 모델  
Fig. 3. Proposed V2I network model

표 1. 본 논문에서 사용된 변수 기호  
Table 1. Notations of variables in this paper

Notation	Description
$BW$	Bandwidth for broadcast channels
$d_0$	Width of the road
$d_{in}$	Distance between the RSUs
$d_h$	Height of the RSU
$\theta$	Half of the angular range covered by the RSU antenna
$v$	Speed of vehicle
$P_B$	Transmit power of RSU
$N_0$	Power spectrum density of the white gaussian noise
$\gamma$	Path loss exponent

$$D_T = \rho \cdot \tan \theta \cdot \log [1 + \xi \cos^\gamma \theta] + \rho \cdot \xi \cdot \gamma \cdot \psi_{\cos \theta}(\gamma, \xi) \quad (4)$$

여기서  $\theta, \rho, \xi, \psi_{\cos \theta}(\gamma, \xi)$ 는 각각 아래와 같다.

$$\theta = \tan^{-1} \left( \frac{d_{in}}{2d_h} \right),$$

$$\rho = \frac{BW \cdot d_0}{v},$$

$$\xi = \left( \frac{P_B}{BW \cdot N_0} \right) \left( \frac{d_0}{2} \right)^{-\gamma},$$

$$\psi_{\cos \theta}(\gamma, \xi) = \int_1^{\cos \theta} \frac{t^{\gamma-2} \sqrt{1-t^2}}{1 + \xi \cdot t^\gamma} dt.$$

정보량 수식을 설명하기 위한 변수들의 표기법은 표 1에서 제시한다. 수식 (4)를 통하여 최대 정보량  $D_T$ 는 OBU의 이동 속도의 관계는  $D_T \propto \rho \propto \frac{1}{v}$  이므로 이동 속도가 증가할수록 최대 정보량은 감소하는 것을 확인할 수 있다. 본 논문에서 V2I 환경의 정보량 감소의 원인은 차량의 속도 변화에 따른 패킷 손실이라고 가정하며 손실률 및 채널의 상태는 일정하다고 가정한다.

#### IV. 문제 정의 및 해결 방법

본 연구는 V2I 환경에서 브로드캐스팅 채널을 통하여 RSU로부터 OBU에 데이터를 공유할 때, 인코딩 비율은 유지하며 데이터 복원율을 높이고자 한다. 차

량의 속도에 따라 전달할 수 있는 최대 정보량이 결정됨에 따라 최대의 데이터를 전송하며 디코딩 성능을 높이는 것이 중요하다. 한편, 브로드캐스팅 채널을 통하여 데이터를 전달할 경우 패킷 손실이 발생하여 복원이 어려운 문제점이 있다. 본 논문에서는 손실된 패킷에 대한 디코딩 어려움을 감소시키는 방법으로 브로드캐스팅 채널에서 SNC를 이용한 데이터 전송 방법을 제안하기 위하여 다음의 최적화 문제를 정의한다.

$$\begin{aligned} & \text{minimize } \sum_i^K \| X_i - \hat{X}(\lambda, D, p)_i \|_0 \\ & \text{subject to } D \leq D_T \end{aligned} \quad (5)$$

여기서  $\lambda$ 는 인코딩 비율로 손실된 패킷을 복원하기 위하여 전송하는 RLNC 패킷을 통해  $\lambda = \frac{N}{N+K}$ 으로 계산된다. 또한,  $\hat{X}(\lambda, D, p)_i$ 는 SNC를 통하여 디코딩된 데이터이며  $\lambda$ 와 전송 가능한 정보량  $D$ , 패킷 손실률  $p$ 에 의하여 결정된다. 본 논문에서 정보량은 최대 정보량에 맞추어 전송한다. 또한, 인코딩 비율은 채널 상태에 따라 결정되나 동일한 채널 상태에 대해서는 인코딩 비율을 유지하며 디코딩 어려움을 감소시키고자 한다. 이를 위하여 회귀 분석 모델을 이용하여 교통 데이터를 기반으로 최대 정보량을 예측한 후, 채널 상태에 따라 RLNC 패킷 데이터를 전송하고자 한다.

#### V. 제안하는 시스템

본 논문에서는 V2I 환경에서 SNC를 이용한 데이터 공유 시스템을 그림 4와 같이 제안한다. RSU는 다수의 OBU와 연결되어있으며 브로드캐스팅을 통하여 데이터를 모든 OBU에 전달한다. RSU는 먼저 OBU의 속도에 따른 정보량을 예측하여 인코딩 넘버를 결

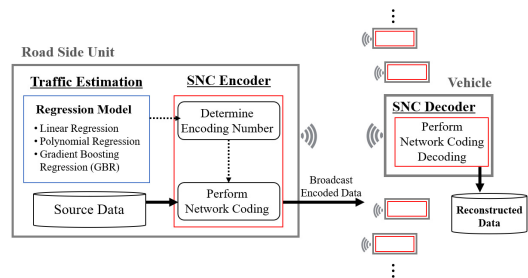


그림 4. 제안하는 V2I 시스템 구조  
Fig. 4. Proposed V2I System Architecture.

정하고 SNC를 이용하여 데이터를 인코딩한다. 데이터를 전달한 이후 OBU에서는 SNC 디코딩 과정을 거쳐 원 데이터를 복원한다.

5.1 이동 속도에 따른 최대 정보량 분석을 위한 회귀 모델 적용

시간에 따라 고속도로에 있는 차량의 평균 속도는 계속해서 변하기 때문에, 전송 가능한 최대 정보량도 지속적으로 변비례하여 변한다. 그러나 고속도로 환경의 차량 속도 및 정보량은 매우 같은 요일에 나타나는 데이터 형태가 매우 비슷한 패턴을 보인다. 따라서 특정 시간에 나타나는 다수의 OBU 데이터 중 브로드캐스팅 환경을 고려하여 정보량의 대푯값을 표현하기 위해, 본 논문에서는 회귀 분석 모델을 이용하고자 한다. 시간에 따라 변하는 다수의 데이터를 대표로 표현하기 위하여 많이 이용되는 회귀 분석 모델을 이동 속도에 따른 최대 정보량 예측 분석에 사용하기 위하여 3가지의 알고리즘을 적용하였다. 첫 번째로 이용한 선형 회귀 분석(linear regression analysis)은 회귀 분석 알고리즘 중 가장 기본이 되는 분석 방법으로 간단하게 적용할 수 있는 장점이 있다<sup>[15]</sup>. 그러나 데이터 포인트가 선형 모델로서 해석이 어려운 패턴에 대해서는 제한된 성능을 보이는 단점이 있다. 이러한 문제점을 해결하기 위하여, 두 번째로는 다항 회귀 분석(polynomial regression analysis) 방법을 본 논문에서는 사용하였다. 회귀 분석 방법으로 많이 사용되는 다항 회귀 분석 알고리즘은 선형 회귀 분석이 해석하지 못하는 구분 형태의 데이터 패턴을 분석하기 위하여 많이 이용되며, 고차 다항식을 통하여 더 낮은 오차를 구할 수 있지만, 과적합을 발생시키는 원인이 되기도 한다<sup>[15]</sup>. 마지막으로 사용한 회귀 분석 방법은 GBR(Gradient Boosting Regression)로 선형 및 다항 회귀 분석 방법과 마찬가지로 널리 이용되는 분석 방법이다<sup>[16]</sup>. GBR(Gradient Boosting Regression)은 의사 결정 트리(decision tree)를 이용하여 개별로 학습된 여러 모델을 결합하여 학습하는 앙상블 약한 학습기(ensemble weak learner)를 통해 데이터의 패턴을 예측한다<sup>[17]</sup>. 따라서, 의사 결정 트리의 개수와 깊이 등의 다양한 변수들이 GBR의 성능을 결정하게 된다. 한편, GBR의 경우 시간에 따른 변화가 빠른 데이터에 대하여 빠른 전환이 가능하여 좋은 성능을 보이는 장점이 있다.

따라서 본 논문에서는 다양한 회귀 분석 방법 중, 선형 회귀 분석, 다항 회귀 분석, GBR을 이용하여 OBU의 최대 정보량을 예측하고자 하며 이를 통하여

SNC의 인코딩 넘버를 결정하고자 한다.

5.2 SNC를 이용한 데이터 전송

차량의 속도에 따라 최대 정보량이 변하기 때문에, RSU는 OBU에 전달할 데이터에 대하여 함께 SNC 인코딩할 데이터의 수, 즉 인코딩 넘버를 결정해야 한다. 수식 (4)에 따라, 시간  $t$ 에 대하여 예측된 평균 속도  $v^*$ 를 가지고 있는 도로 환경에서 최대로 전송될 수 있는 최적 정보량을  $D_T^*(t)$ 라고 하자. 채널 손실률이  $p$ 일 때, 원본 데이터 집합  $S$  중 인코딩 넘버  $K^*(t)$ 개의 데이터  $X$ 를 인코딩하기 위해서는 아래와 같은 방법을 통하여 인코딩 넘버를 결정한다.

$$K^*(t) = D_T^*(t) \times (1-p) \tag{6}$$

즉, 인코딩 넘버는  $K^*(t)$ 는 최대 정보량 중에서 손실 없이 받을 수 있는 데이터의 양으로 설정한다. 인코딩 넘버가 결정된 후에는 수식 (2)와 비슷하게 아래와 같이 선형 결합을 통하여 인코딩된 데이터  $Y^*$ 를 계산한다.

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_K \\ Y_{K'+1} \\ \vdots \\ Y_{K'+N'} \end{bmatrix} = \begin{bmatrix} I_{K'} \\ c_{11} \cdots c_{1K'} \\ \vdots \\ c_{N'1} \cdots c_{N'K'} \end{bmatrix} \otimes \begin{bmatrix} X_1 \\ \vdots \\ X_{K'} \end{bmatrix} \tag{7}$$

여기서,  $I_{K'}$ 는  $K^* \times K^*$  항등행렬이며 코딩 계수 행렬  $C$ 는  $N^* \times K^*$  크기의  $GF(2^M)$ 에서 랜덤으로 선택된 행렬이다. 수식 (7)의 결과로 총  $K^* + N^*$ 개의 인코딩 된 데이터가 RSU에서 OBU로 브로드캐스팅 된다. 한편, 최대 정보량은  $D_T^*(t)$ 로 정해졌기 때문에  $N^* = p \times D_T^*(t)$ 임을 알 수 있다.

RSU에서 전송된 데이터를 OBU에서 수신하여 OBU는 SNC 디코딩 과정을 수행하게 된다. 디코딩은 (3)과 마찬가지로 역행렬 연산을 통하여 계산하게 되는데 기존 방식과 마찬가지로 전체 수신된 데이터의 수가  $\widetilde{N}^* + \widetilde{K}^* < K^*$ 일 경우 코딩 계수 행렬의 상굴러리티 때문에 복원이 불가능하여 전송된 데이터 집합 모두 잃게 된다. 그러나  $\widetilde{N}^* + \widetilde{K}^* \geq K^*$ 일 경우, 코딩 계수가 항등 행렬의 일부인 데이터 패킷을 우선순위로 하여  $\widetilde{N}^* + \widetilde{K}^*$ 개의 데이터 패킷 중  $K^*$  패킷에 대

하여 디코딩 과정을 수행한다.

다음 장에서는 제안된 알고리즘에 대하여 실제 데이터를 이용하여 실험하고 성능을 평가하고자 한다.

### VI. 실험

본 논문에서는 교통 데이터를 이용하여 제안된 알고리즘의 성능을 확인하고 분석하고자 한다. 제안된 알고리즘의 성능을 평가하기 위하여 본 논문에서는 고속도로 차량의 평균 속도 데이터를 사용하였다<sup>[8]</sup>. 차량의 평균 속도는 5분 간격으로 수집되었으며 2018년 1월 1일부터 2018년 4월 30일까지의 목요일 데이터를 사용하였다. 한편, 본 논문에서는 전체의 데이터 중 75%의 데이터가 학습 데이터로 사용하였고 나머지 25%는 테스트 데이터로 사용되었다.

본 실험에서의 환경은 RSU와 OBU가 존재하는 V2I 환경을 반영하여 다음과 같이 설정하였다. 본 논문에서는 IEEE 802.11p/DSRC EDCA 통신 기술에 따라 5.9GHz 주파수를 이용하여 RSU에서 OBU로 256 Byte의 데이터를 지속적으로 전달한다. 이때, RSU에서 OBU로 전달하는 채널의 BW는 10MHz로 설정하였으며, 도로의 너비는  $d_0 = 20m$ , RSU 사이의 거리는 VDS(Vehicle Detection System) 측정 거리에 따라  $d_{\in} = 1km$ , RSU의 높이는  $d_h = 12m$ 로 설정하였다. 또한, RSU의 송신 출력  $P_B = 1W$ 이며, 열잡음 출력은  $N_0 = 3.34 \times 10^{-9} W$ 로 설정하였다<sup>[13]</sup>. 그리고, 경로 손실 상수  $\gamma = 2$ , 패킷 손실률은  $p \sim N(0.2, 0.1^2)$ 로 설정하여 실험을 진행하였다. 한편, OBU는 고속도로에서 주행하고 있기 때문에 차량의 이동 패턴은 한 방향으로 일관되게 진행한다고 가정하였다. 본 실험의 성능 분석을 위하여 SNC를 사용하지 않은 브로드캐스팅 전송 방법(without SNC), 정보량을 예측하지 않고 데이터를 전송하는 SNC(existing SNC), 그리고 각 회귀 모델에 따라 정보량을 예측하여 데이터를 전달한 제안된 SNC 알고리즘을 비교하였다. 모든 SNC 인코딩 및 디코딩 과정에서 사용된 GF의 사이즈는  $2^8$ 으로 동일하였다.

본 실험은 신호등이 존재하는 복잡한 도로 구조와 차선 변경이 빈번히 일어나는 도로가 아닌 고속도로 환경을 고려하기 때문에 NetSim, SUMO 등의 V2X 네트워크 시뮬레이션은 불필요하여<sup>[8]</sup>, IEEE 802.11p/DSRC EDCA 기술을 이용한 네트워크 코딩 성능 검증을 Intel Core i7 3.40GHz CPU, 8.00GB RAM, Windows 10 시스템에서 MATLAB R2019a

버전에서 진행하였다. 그러나 차량의 속도 데이터를 전송량으로 변환하고 이에 대한 대표 전송량 값을 구하기 위하여 회귀 분석 알고리즘은 python 3.7.3 환경에서 진행하였으며, 수집된 데이터에 대하여 MATLAB으로 디코딩 성능 검증을 하였다.

본 논문에서 사용된 회귀 분석 중, 다항 회귀 분석은 정보량 데이터 패턴에 따라 6차 함수를 사용하였으며, GBR의 경우 결정 트리의 수는 15, 최대 심도는 3, 러닝 레이트는 0.3, leaf 노드의 최소 샘플 수는 5, 불순도 MSE(Mean Square Error), 변위치는 0.9로 설정하여 실험을 진행하였다.

그림 5는 회귀 분석 모델을 이용하여 정보량을 예측하고 SNC를 이용하여 데이터를 전송한 그래프를 나타낸다. 표 2는 테스트 데이터와 각 알고리즘의 정보량 예측 간의 MSE, 각 알고리즘의 인코딩 비율, 그

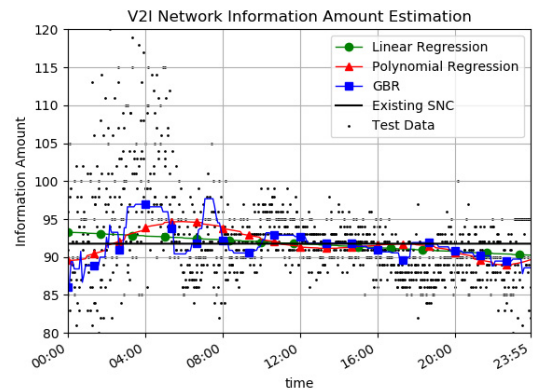


그림 5. 교통 데이터를 이용한 정보량 예측 실험 결과  
Fig. 5. Experimental result of information amount estimation using traffic data.

표 2. 제안된 시스템의 성능  
Table 2. Performance of the proposed system

	MSE	Encoding Rate(%)	Decoding Performance(%)
Without SNC	34.69	0	78.59
Existing SNC	34.69	78.24	90.25
Linear Regression -SNC	31.33	78.79	92.98
Polynomial Regression -SNC	32.10	78.82	93.06
GBR -SNC	28.95	78.70	93.31



리고 디코딩 성능을 나타낸다. 그림 5와 표 2의 MSE를 통하여 확인할 수 있는 것과 같이 회귀 모델을 이용하여 데이터를 전송할 경우 기존 방법과 비교하여 정보량 오차율이 감소하는 것을 확인할 수 있다. 또한, 정보량의 변화율이 시간에 따라 빠르게 때문에 GBR이 MSE 28.95로 더 정확하게 예측함을 확인할 수 있다.

표 2의 디코딩 비율은 수식 (5)에서 계산한 값을 통하여 확인할 수 있다. 회귀 모델을 이용하여 데이터의 최대 정보량을 예측한 후 데이터를 전달할 경우 인코딩 비율  $\lambda$ 는 유지를 하면서 데이터의 디코딩 성능을 증가시킴을 확인할 수 있다. 특히 어떠한 인코딩 없이 데이터를 브로드캐스팅 채널을 통하여 전송하였을 경우 78.59%로 낮은 디코딩 성능을 보이는 것을 알 수 있으며, 기존 정보량 예측 없이 SNC를 이용하여 데이터를 전송하였을 경우 90.25%의 성능을 보임을 확인할 수 있다. 그러나, 제안된 SNC 방법을 통하여 최대 93.91%까지 성능을 향상시킬 수 있음을 실험을 통하여 확인할 수 있다.

따라서 V2I 환경에서 회귀 모델을 이용하여 최대 정보량을 예측한 후, SNC를 기반으로 데이터를 전송할 경우 인코딩 비율은 유지하면서 디코딩 성능은 높아짐을 알 수 있다.

## VII. 결 론

본 논문에서는 V2I 환경에서 교통 데이터 기반 SNC를 이용한 데이터 전송 방법에 대하여 제안하였다. 제안하는 시스템은 RSU에서 OBU로 데이터를 공유하고자 할 때, IEEE 802.11p/DSRC EDCA 브로드캐스팅 채널을 통하여 데이터를 전달한다. 그러나 빠르게 이동하는 OBU의 속도에 따라 RSU에 연결되는 시간이 결정되게 되기 때문에, 최대로 전송할 수 있는 정보량은 시간에 따라 계속해서 변하는 특징이 있다. 더욱이 무선 브로드캐스팅 채널의 경우 패킷 손실이 발생하는 문제점이 있어 데이터 복원이 어려운 문제점이 있다. 따라서 본 논문에서는 수집이 어려웠던 V2I 네트워크 데이터를 도로 환경 데이터를 이용하여 최대 전송 가능한 정보량 데이터를 간접적으로 생성하였으며, 회귀 분석 모델을 통하여 시간에 따라 변화하는 최대 정보량을 표현하는 대표 최적 정보량을 얻었다. 이후, 최적의 인코딩 넘버를 결정하고 IEEE 802.11p/DSRC EDCA 시스템에서 SNC를 이용하여 데이터를 전달하였다. 그 결과, 예측된 정보량과 채널의 상태에 따라 일정 비율의 RLNC 패킷을 전달함으

로써 손실된 패킷에 대한 디코딩이 성과 채널 사용 효율성을 높였으며, 이를 실험을 통해 확인하였다.

## References

- [1] R. Hosh, R. Pragathi, S. Ullas, and S. Borra, "Intelligent transportation systems: A survey," *2017 Int. Conf. Cir., Contr., and Commun.*, pp. 160-165, Dec. 2010.
- [2] K. Dar, M. Bakhouya, J. Gaber, and M. Wack, "Wireless communication technologies for ITS applications," *IEEE Commun. Mag.*, vol. 48, no. 5, pp. 159-162, 2010.
- [3] J. Kim and S. Lee, "Reliable routing protocol for vehicle to infrastructure communications in VANET," *J. KICS*, vol. 34, no. 8, pp. 839-845, 2009.
- [4] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," *IEEE Int. Conf. Computer Commun.*, vol. 2, pp. 1404-1413, 2004.
- [5] R. Verdone, "Multihop R-ALOHA for intervehicle communications at millimeter waves," *IEEE Trans. Veh. Technol.*, vol. 46, no. 4, pp. 992-1005, Nov. 1997.
- [6] IEEE 802.11e/D4.4, *Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, IEEE 802.11e/D4.4, Jun. 2003.
- [7] D. E. Lucani, M. Médard, and M. Stojanovic, "Systematic network coding for time-division duplexing," *IEEE Int. Symp. Information Theory*, pp. 2403-2407, 2010.
- [8] Q. Wang, P. Fan, and K. B. Letaief, "On the joint V2I and V2V scheduling for cooperative VANETs with network coding," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 62-73, 2012.
- [9] V. Vijayakumar and K. S. Joseph, "Adaptive load balancing schema for efficient data dissemination in vehicular Ad-Hoc network VANET," *Alexandria Eng. J.*, pp. 1-29, Jun. 2019.

[10] M. Kwon, S. Park, and H. Park, "Analysis on decoding time for systematic network coding over lossy channel," in *Proc. Symp. KICS*, pp. 1155-1156, Jun. 2016.

[11] J. Kwon, L. Moon, and H. Park, "Systematic network coding based high fidelity RLNC encoding number in V2I networks," *Joint Conf. Commun. and Inf.*, pp. 234-235, Apr. 2018.

[12] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Information Theory*, vol. 46, no. 4, pp. 1204-1216, Jul. 2000.

[13] D. Nguyen, T. Tran, T. Nguyen, and B. Bose, "Wireless broadcast using network coding," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 914-925, 2008.

[14] J. Heide, M. V. Pedersen, F. H. Fitzek, and T. Larsen, "Network coding for mobile devices-systematic binary random rateless codes," *IEEE Int. Conf. Commun. Workshops*, pp. 1-6, 2009.

[15] M. Kwon and H. Park, "Analysis on decoding error rate of systematic network coding," *IEEE Int. Conf. Consumer Electronics*, 2017.

[16] D. C. Montgomery, E. A. Peck, and G. G. Vining, "Introduction to linear regression analysis," John Wiley & Sons, vol. 821, pp. 11-20, 2012.

[17] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308-324, 2015.

[18] *Korea highway system, Korea Expressway Corporation*(2018), Retrieved Jul. 15, 2019, from: <http://data.ex.co.kr/portal/fdwn/view?type=VD S&num=38&requestfrom=dataset>.

**권 정 민 (Jungmin Kwon)**



2015년 2월 : 이화여자대학교  
전자공학과 졸업  
2017년 2월 : 이화여자대학교  
전자공학과 석사과정 졸업  
2017년 3월~현재: 이화여자대  
학교 전자전기공학과 박사과  
정

<관심분야> 네트워크 코딩, 머신러닝 기반 네트워  
크 자동화 시스템 설계

[ORCID:0000-0002-9822-3047]

**박 형 곤 (Hyunggon Park)**



2004년 2월 : 포항공과대학교 전  
자전기공학과 졸업  
2006년 3월 : University of  
California, Los Angeles  
(UCLA) M.S.  
2008년 12월 : University of  
California, Los Angeles  
(UCLA) Ph.D.

2010년~현재 : 이화여자대학교 전자전기공학과 부교  
수

<관심분야> 멀티에이전트 네트워크 시스템, 머신러닝  
기반 분산적 의사 결정 전략, 게임이론 기반 네트워  
크 분산적 자원 관리, 네트워크 코딩

[ORCID:0000-0002-5079-1504]