

Research Assessment Exercise 2020
Impact Case Study

University: The Chinese University of Hong Kong

Unit of Assessment (UoA): 12 Electrical & Electronic Engineering

Title of case study: Network Coding is Changing the Landscape of Network Communications

(1) Summary of the impact (indicative maximum 100 words)

Network Coding is an established field of interdisciplinary research in Information Sciences co-founded by CUHK researchers in the late 1990s. Technologies based on network coding have had major applications in cloud storage (Apache Hadoop 3.0 and Ceph) and smart cities (smart lampposts in Hong Kong), in both industry and the scientific community. There have been a number of startup companies making network coding products. Emerging network coding applications include V2X, satellite and deep space communications, underwater acoustic communication, and powerline communication. Adoption of network coding is being considered by various standard organizations.

(2) Underpinning research (indicative maximum 500 words)

The field of Network Coding was co-founded by CUHK researchers (Ning Cai, S.-Y. Robert Li, Raymond W. Yeung) in [3.1] and [3.2] published in 2000 and 2003, respectively. For decades, information was regarded as a commodity in network communication, and data packets are routed from the source to the destination through the communication network very much like parcels being routed in the postal network. In [3.1], the problem of multicasting a single information source in a network was studied. A very simple example, now widely known as the *butterfly network*, shows explicitly the benefit of network coding over routing. This refuted the folklore that well-compressed information bits behave like a commodity, an assumption computer network designs had been based on. The term “network coding” was coined, and the fundamental max-flow min-cut theorem of network coding was proved.

In [3.2], it was proved that the multicast capacity obtained in [3.1] can be achieved by linear network coding. Linear network codes, being highly structured, have advantage over nonlinear network codes in terms of coding complexity and storage requirement. The sufficiency of linear network codes make it possible to implement network coding in real systems.

Network coding theory has been applied to many different domains. In [3.3], an error control theory based on network coding that generalizes the classical algebraic coding theory (which had been studied for over half a century) was developed. Network generalizations of fundamental coding bounds were proved. In a similar spirit, a security theory based on network coding that generalizes the classical problem of secret sharing in cryptography (1979) and its main results was developed in [3.4].

All the above works consider network coding operations at the digital level. In 2006, CUHK researcher Soung Liew and his team launched a new research area called Physical-layer Network Coding (PNC) [3.5] by exploiting the network coding operation that occurs in nature when electromagnetic (EM) waves are superimposed on one another. In the two-way relay setup where the two users can transmit to the relay simultaneously, the employment of PNC can boost the throughput by 100% at high signal-to-noise ratio.

In [3.5], CUHK researcher Raymond Yeung and his postdoc Shenghao Yang invented BATCHed Sparse Codes (BATS codes), the first efficient implementation of linear network coding. It is well known that transmission on a wireless multi-hop network cannot sustain more than 5 or 6 hops due to accumulation of packet loss. BATS code solves this longstanding problem in wireless communication. With BATS code, transmission can sustain tens or even hundreds of hops. BATS code has already been implemented in the Hong Kong Government’s pilot smart lamppost system.

Other potential applications of BATS code include 5G, V2X, satellite networks, underwater acoustic communication, etc.

N. Cai, S.-Y.R. Li, and R.W. Yeung received the 2016 IEEE Eric E. Sumner Award for their pioneering contributions to Network Coding. Previously, [3.2] received the 2005 IEEE Information Theory Society Paper Award, and subsequently [3.1] received the 2018 ACM SIGMOBILE Test-of-Time Paper Award.]

(3) References to the research (indicative maximum of 6 references)

- [3.1] [R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung (in alphabetical order), “Network information flow,” *IEEE Transactions on Information Theory*, IT-46, pp. 1204-1216, July 2000. **Google Scholar citation: 9,325.**
- [3.2] S.-Y. R. Li, R. W. Yeung and N. Cai, “Linear network coding,” *IEEE Transactions on Information Theory*, IT-49, pp. 371-381, Feb 2003. **Google Scholar citation: 4,009.**
- [3.3] R. W. Yeung and N. Cai, “Network error correction, Parts I & II,” *Communications in Information and Systems*, 6: 19–54, 2006.
- [3.4] N. Cai and R. W. Yeung, “Secure Network Coding on a Wiretap Network,” *IEEE Transactions on Information Theory*, IT-57, pp. 424-435, Jan 2011.
- [3.5] S. Zhang, S. C. Liew, and P. P. Lam, “Physical-layer network coding,” The 12th annual international conference on Mobile computing and networking (MobiCom '06), Los Angeles, CA, Sept 23-29, 2006. Google Scholar citation: 2,130.
- [3.6] S. Yang and R. W. Yeung, “Batched Sparse Codes,” *IEEE Transactions on Information Theory*, IT-60, pp. 5322- 5346, July 2014.]

(4) Details of the impact (indicative maximum 750 words)

Network coding is finding numerous applications in computer networks, wireless communication, distributed data storage, satellite and deep space communications, underwater acoustic communication, and powerline communication. Each such application is changing the related industry. There are a number of startup companies making network coding products. Seven (7) Internet Drafts (potentially leading to Internet standards) on network coding applications have been submitted to Internet Engineering Task Force (IETF), including one on BATS code by CUHK researcher Raymond Yeung and his collaborators.

<https://datatracker.ietf.org/rg/nwcrp/documents/> (2018-19)

Cloud Storage

An important application of network coding is proposed in the seminal paper

G. Dimakis et al., “**Network Coding** for Distributed Storage Systems,” *IEEE Transactions on Information Theory*, IT-56, pp. 4539- 4551, Sept 2010 (Google Scholar citation: 1,768), where **regenerating codes** were proposed for distributed data storage. Compared with conventional storage codes, regenerating codes require a smaller amount of data to be downloaded when repairing a failed node. In a reasonable setting, the saving can be over 60%.

CUHK has also significantly contributed to the application of regenerating codes:

1. In 2011, CUHK researcher Patrick Lee and his team built the world’s first data storage system based on regenerating code. The system has been deployed at the CUHK Information Technology Service Centre for storing data sets for the Daya Bay Reactor Neutrino Experiment (2012–present), one of the top 10 scientific breakthroughs in 2012 selected by *Science*:

<https://science.sciencemag.org/content/338/6114/1525.full#sec-3>

https://en.wikipedia.org/wiki/Daya_Bay_Reactor_Neutrino_Experiment

This system has about 1 PByte storage size and has been running as a CERN (European Organization for Nuclear Research) ATLAS Tier-2 site since Jan 2018, under the French Cloud of computing sites. We are working on extending such for various CERN projects, pending further test with genuine data of CERN.

2. Si-Tech Information Technology Ltd. is a Beijing based company that has been focusing on mobile billing software development and operation in Mainland China for over two decades. The number of employees exceeds 2,000. The company was IPOed in February 2017 at the Growth Enterprise Board of Shenzhen Stock Exchange (stock code: 300608). Through a collaboration with Si-Tech, file systems based on regenerating codes have already been deployed at the Si-Tech data centers, serving China Unicom customers in the Guangdong and Sichuan provinces in real time. Similar file systems will significantly enhance all Si-Tech's existing file systems, currently serving China Unicom, China Telecom, and China Mobile customers in 12 application scenarios. **Here Data Technology**, a startup company resulting from this collaboration, was set up in Shenzhen in 2017.

There have also been other major applications of regenerating codes:

1. *Piggyback codes* have been incorporated into Apache Hadoop 3.0 (2017), a widely used big-data storage and analytics system with hosting in the cloud by Amazon, Google, IBM, Microsoft, Oracle, SAP, etc.
https://en.wikipedia.org/wiki/Apache_Hadoop
2. *Clay codes* have been incorporated into the Ceph (2018), a widely used distributed storage system developed by CERN, Cisco, Fujitsu, Intel, SanDisk, etc.
[https://en.wikipedia.org/wiki/Ceph_\(software\)](https://en.wikipedia.org/wiki/Ceph_(software))

Smart Lampposts

Smart lampposts are the key infrastructure of smart cities. Each smart lamppost may be equipped with a video camera for monitoring traffic and pedestrian flows, sensors for monitoring air quality and weather conditions, an access point for providing Wi-Fi services, etc. The equipment on these lampposts must be connected to the Internet backbone. Realistically, it is possible to provide optical fibre connection to only a small number of lampposts. With BATS code, the other lampposts can be connected to these lampposts by forming a wireless multi-hop network.

In the Hong Kong Government's pilot smart lamppost system, the communication infrastructure for lampposts with no direct optical fibre connection will be enabled by BATS code. Initially 400 smart lampposts will be installed. After the pilot phase, several tens of thousands of smart lampposts will be deployed throughout the city. **n-hop technologies**, a startup company set up in Hong Kong by CUHK researcher Raymond Yeung, has been commissioned by the Hong Kong Government to provide the BATS code technology in the project. The technology has already been successfully deployed at 3 locations in East Kowloon (a total of 36 smart lampposts).

CUHK Startups on Network Coding

Here Data Technology, Shenzhen (since 2017)

Technology: network coding data storage

Initial investment: RMB 20M

No. of employees: 24

CU Coding, Hong Kong (since 2018)

Technologies: network coding data storage and physical-layer network coding

Initial investment: HKD 20M

No. of employees: 15

<http://www.cucoding.cn/aboutus/>

n-hop technologies, Hong Kong (since 2018)

Technology: BATS code

Initial investment: HKD 2.2M

No. of employees: 3

<http://www.n-hop.com>

Other Active Companies/Startups on Network Coding (all related to MIT)

<http://codeontechnologies.com>

<http://steinwurf.com>

<http://www.chocolate-cloud.cc> |

(5) Sources to corroborate the impact (indicative maximum of 10 references)

(1) Seven (7) Internet Drafts on network coding applications submitted to Internet Engineering Task Force (IETF)

<https://datatracker.ietf.org/rg/nwcrp/documents/>

(2-5) Applications of Network Coding in Distributed Data Storage

2. The seminal paper that applied network coding to distributed data storage and proposed regenerating codes

G. Dimakis et al., “Network Coding for Distributed Storage Systems,” *IEEE Transactions on Information Theory*, IT-56, pp. 4539- 4551, Sept 2010. Google Scholar citation: 1,768.

3. Application in Data Centers of Si-Tech Information Technology Ltd.

See attached letter from Here Data Technology.

4. Application of Piggyback Codes (a class of regenerating codes) in Apache Hadoop

K. V. Rashmi et al., “A Piggybacking Design Framework for Read- and Download-efficient Distributed Storage Codes,” *IEEE Transactions on Information Theory*, IT-63, pp. 5802-5820, Sept. 2017.

5. Application of Clay Codes (a class of regenerating codes) in Ceph

M. Vajha et al., “Clay codes: Moulding MDS Codes to Yield an MSR Code,” USENIX FAST 2018.

(6) Application of BATS Codes in the Hong Kong Government’s Smart Lamppost System

See attached letter from the Government Chief Information Officer, Hong Kong Special Administrative Region.

(7) Report on network coding in MIT Technology Review, October 2015

<https://www.technologyreview.com/s/542131/signal-intelligence/> |