NC-CELL: Network Coding-based Content Distribution in Cellular Networks for Cloud Applications

Claudio Fiandrino, Dzmitry Kliazovich, Pascal Bouvry
University of Luxembourg
6, Rue Richard Coudenhove Kalergi, Luxembourg
{name.surname}@uni.lu

Albert Y. Zomaya
University of Sydney
Sidney, NSW 2006, Australia
albert.zomaya@sydney.edu.au

Abstract—The popularity of cloud applications surged in the last years. Billions of mobile devices remain always connected. Location services, online games, social networking and navigation are a just few examples of "always on" cloud applications in which the same or partially overlapping content is delivered to multiple users. In this paper, we propose a technique, called NC-CELL, which uses network coding to foster content distribution in mobile cellular networks. Specifically, NC-CELL implements a software module at mobile base stations (or eNodeBs) which scans in transit traffic and looks for opportunities to code packets destined to different mobile users together. The proposed approach can significantly improve cell throughput and is particularly relevant for delay tolerant content distribution.

Index Terms—Network coding, cellular networks, content distribution.

I. INTRODUCTION

Mobile cloud applications is one of the fastest growing markets in the view of the wide spread of new smartphones, tablets and netbooks which feature advanced computing, broadband access and storage capabilities. With more than 7 billion mobile users [1], the ubiquity and mobility are certainly two main features leading the raise of popularity of mobile cloud applications [2]. By 2017 there will be 4.4 billion people using mobile applications [1] and it will correspond to $45 billion market according to Visiongain [3].

A large number of popular mobile cloud applications, such as navigation, location-aware services, online games and multimedia streaming, often deliver the same pieces of content to multiple users. Many of these applications, for example Google Now [4], run most of the time in background autonomously without requiring an explicit input from users. Then, when geographically co-located user terminals ask for the same or partially overlapping content, the network accomplishes these requests separately. As a consequence, it uses a large number of available resources. Being co-located, the users are likely to be served by the same base station and, instead of sending several data flows with the same content on different channels, these flows can be combined at the base station and broadcast to the interested users using just a fraction of available network resources.
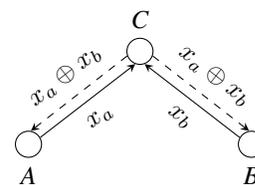

Fig. 1. Basic network coding example.

In this paper we investigate if network coding can be used to improve content distribution in mobile cellular networks. Network coding is proved to have a potential to increase network throughput considerably by leveraging shared nature of the broadcast wireless medium [5]. To illustrate, consider the example in Fig. 1. Node A needs to transmit packet x_a to node B and node B needs to transmit packet x_b to A. Being located in the same cell, nodes A and B are required to communicate through the same base station node C. As soon as x_a and x_b are received at node C, both packets are combined into one coded packet using a linear operation on packets' data, for example XOR [5]. Then the coded packet $x_a \oplus x_b$ is transmitted by node C in broadcast and successfully reaches nodes A and B. Nodes A and B can decode x_b and x_a respectively by performing XOR with previously sent packets: $(x_a \oplus x_b) \oplus x_a$ to decode x_b and $(x_a \oplus x_b) \oplus x_b$ to decode x_a. As a result, only 3 transmissions are needed in total instead of 4 to deliver both packets.

Network coding was originally proposed for improving throughput of multicast communications [6]. The key idea is to allow intermediate nodes not only perform forwarding operations, but also to create new packets by combining two or more packets from the incoming flows. Throughput enhancement is just one of the possible applications of network coding. Robustness, network tomography, security and storage are several other examples [7]. In wireless and cellular networks, network coding is typically applied to improve throughput in scenarios with relays, when base station transmissions are relayed at least once before they can reach end users, and robustness [8]–[11]. However, even though relays are excellent points to perform network coding, they are not very common in current cellular networks. For this, we apply network coding in
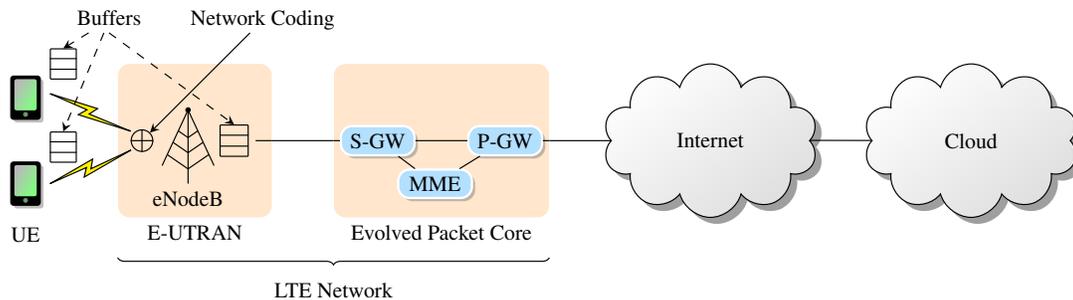
Fig. 2. System scenario.

cellular networks with no relays with a purpose to improve cell throughput capacity for unicast traffic with partially overlapping content.

In this paper we propose a network coding based technique, called NC-CELL (Network Coding in CELLular networks), which improves efficiency of content distribution for cloud applications in mobile cellular networks to geographically co-located users. NC-CELL allows cellular base stations to monitor in transit traffic, caching it locally. Then, whenever a coding opportunity is detected, which happens when the base station is able to distribute information needed by two or more users with a single transmission, a coded packet is created and broadcast in the cell.

The rest of the paper is organized as follows. Section II discusses related work. Section III presents motivation and describes the proposed technique for network coding based content distribution. Performance evaluation and results are presented in Section IV. Finally, Section V concludes the work and outlines directions for future research on the topic.

II. RELATED WORK

In this section we review the background on network coding by discussing the basics and then reviewing network coding application in cellular networks as well as in cloud-based scenarios.

Network coding basics: The research on network coding started with the pioneering work [6] which demonstrated that the performance of multicast communications can be increased if routers are allowed to combine packets and information flows. Linear network coding (LNC) was introduced later [12]. LNC allows routers to create linear combinations of the incoming packets, which appears to be sufficient to achieve multicast capacity limit. XOR coding, which is a special case of LNC, was first introduced for wireless network [5]. It demonstrated significant throughput improvement and appeared to be the most useful for unicast flows.

Network coding in cellular networks: In mobile cellular networks, network coding is mainly used to improve throughput in scenarios with relays and to enhance error recovery techniques. For error correction, an XOR coding enabled HARQ can be used [9], [10]. Reference [9] exploits radio relay nodes, such as low power eNodeBs introduced in LTE-A, to combine flows for enhancing both coverage and network capacity. To improve efficiency of error recovery, network nodes can use random

coefficients taken from a sufficiently large Galois field [11]. This approach, called Random Linear Network Coding (RLNC), demonstrates excellent throughput performance in decentralized network scenarios. In LTE, RLNC can be used to enable energy-aware communications [13] and error recovery [14]. RLNC is typically implemented at the LTE MAC layer and often requires cooperation between mobile nodes [15].

Network coding in cloud-based scenarios: In wired networks and cloud data centers network coding can be used to improve redundancy and reliability of distributed storage systems [16], [17] by increasing fault tolerance. With a focus on cloud-based multimedia applications, network coding can optimize power consumption of co-located user terminals [18] and improve cooperative download experience of mobile users [19]. Furthermore, stochastic analysis confirms network coding ability to enhance efficiency of epidemic routing for mobile cloud networks [20], allowing nodes to opportunistically share their computing or communication resources.

III. NETWORK CODING IN CELLULAR NETWORKS

This section presents a network coding based technique for efficient content distribution for cloud applications in mobile cellular networks. It allows base stations (eNodeBs according to LTE terminology) to monitor in transit traffic, caching it locally, and identify possible coding opportunities. Coding opportunities occur when eNodeB can deliver information to at least two mobile users with a single transmission.

A. Motivation and Scenario Description

Fig. 2 presents reference cellular network scenario. Mobile users (on the left) are connected through LTE network [21] and communicate with cloud applications (on the right). eNodeBs form Evolved UMTS Terrestrial Radio Access Network (E-UTRAN) and are responsible for control and resource management in radio access network. Previous cellular systems were based on circuit-switching, but LTE supports full end-to-end IP connectivity.

In urban environments, standard LTE cells are typically smaller than 1 km in diameter [22], while pico and femto cells are in the order of 500 and 10 meters respectively [23]. These small cell sizes make it very likely that mobile users served by the same eNodeB are located in the same geographical area and often receive completely or partially the same content. Indeed, an increasing number of cloud applications relies on
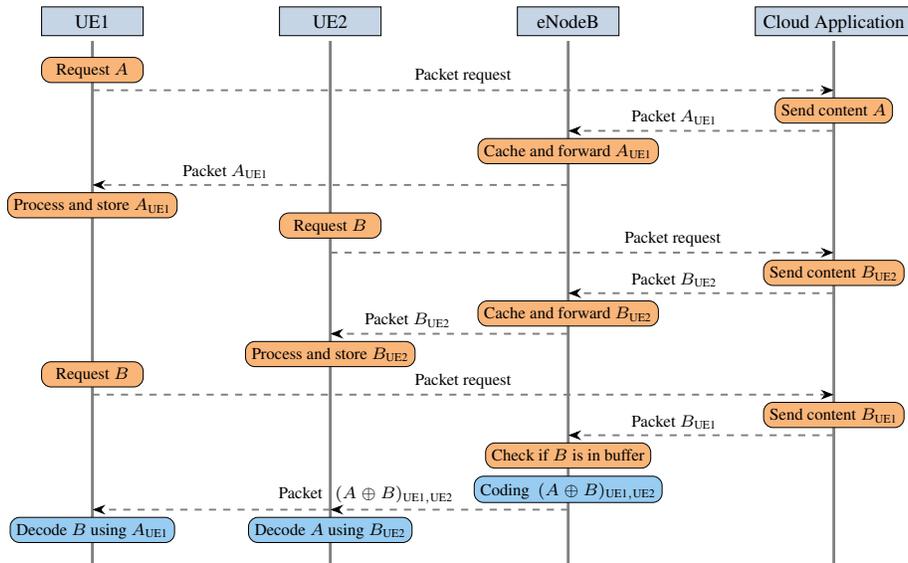
Fig. 3. Network coding with two users.

location-based information or information related to user local environment, while geographically co-located users often have similarities [24]. This information is obtained automatically and on a periodic basis, without requiring an explicit feedback from the user.

For example, Google Now [4] service, which is by default running on all Android devices in background, tries to anticipate users' needs by displaying series of cards created based on the information about users' daily routine, location patterns, information about nearest businesses and restaurants, weather forecasts, road traffic conditions, busses and trains, and other location-based services. Local advertisement applications, such as myCloudLOCAL [25], provide offers according to the user geographical position in an interactive manner. Finally, cars driving through the same neighborhood would need to receive the same part of map, points of interest and driving directions. It is important to note that in the aforementioned scenarios mobile users remain unaware of their neighbors' behavior and intentions for downloading content, which does not permit the use of cooperative solutions [18], [19].

To optimize the delivery of individual flows with partially overlapping content in cellular networks we propose to use a network coding based NC-CELL approach. NC-CELL allows eNodeBs to monitor network traffic destined to the cloud users, cache it locally, and identify eventual coding opportunities. Coding opportunity appears when eNodeB is able to distribute information needed by two or more users with a single transmission of a coded packet.

Packets are combined together using XOR operation applied on packets' data. With respect to traditional LNC techniques, XOR coding is lightweight and 6-10 times faster [26]. In addition, XOR coding is effective in terms of energy consumption [27], which is very important for battery powered devices. In comparison with RLNC, XOR coding allows smaller flexibility in combining packets and has higher latency, but it is more suitable for scenarios with only one node performing encoding operations [28].

Fig. 3 illustrates NC-CELL network coding performed for two users. First, users UE1 and UE2 request and obtain content A and B respectively from the cloud. While content A and content B are passing through eNodeB, they are cached. Whenever any of the users request a content already received by the other user, in this case user UE1 requests content B, which was already delivered to user UE2, a coding opportunity arises. The eNodeB uses this opportunity to encode content A and content B together and send the obtained coded packet $(A \oplus B)_{UE1,UE2}$ to users UE1 and UE2 in broadcast. Having received $(A \oplus B)_{UE1,UE2}$, each user performs decoding by subtracting (applying XOR) previously received content, i.e., user UE1 subtracts content A and user UE2 subtracts content B, for successful decoding.

An important detail of the proposed approach is in the way coded packets are sent. They are delivered in Physical Downlink Shared CHannel (PDSCH) assigned to the primary user. The primary user is a user which makes the content request that leads to coding opportunity, UE1 in our example. The transmission of the coded packet $(A \oplus B)_{UE1,UE2}$ is performed in the same way as other unicast transmissions destined to UE1 and it is subject to error control and HARQ procedures. The transmission bitrate is selected according to the feedback from Channel Quality Indicator (CQI). Secondary users receive coded packets by listening PDSCH channel of the primary user. The information about which channel to listen is broadcast by eNodeB on the Physical Downlink Control CHannel (PDCCH) to all the secondary users. Unlike in the case of primary users, no error correction or HARQ is performed for the secondary users, which decreases probability of successful reception of coded packets for secondary users. However, it is acceptable, as coded packets are just a way to receive additional content which was not explicitly requested by secondary users. Encoding and decoding operations are performed over IP packets and not at the MAC level. This is required to allow eNodeBs analyze
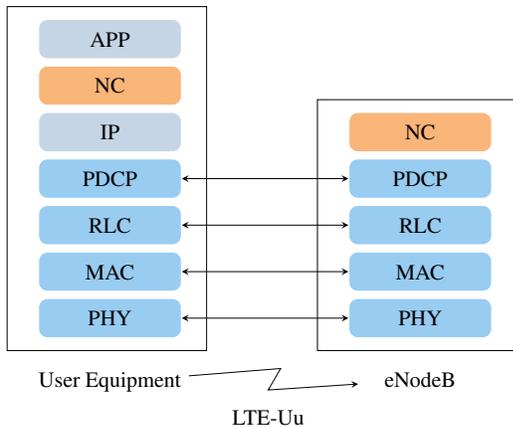
Fig. 4. User plane protocol stack.

packets for encoding before they are passed to the Packet Data Convergence Protocol (PDCP) protocol. Fig. 4 shows the position of network coding (NC) layer in user plane protocol stack.

Only content parts of the packets are combined during coding, while both network and transport layer headers remain that of the primary user. IP optional fields are used to signals IDs of packets that were encoded.

Decoding is performed at client side. For every coded packet received, it is first checked if this packet can be decoded immediately, which requires having one of the two packets used during encoding already received and stored in the buffer. Otherwise, the packet is stored for later decoding. Matches are identified using content IDs carried in IP options field of the coded packet. Each stored packet is assigned with a lifetime timer. As soon as this timer expires the packet is removed from the buffer preventing an overflow. Fig. 5 outlines the details of the described decoding procedure.

B. System Model

The system is modeled as a set of k users $U = \{u_1, u_2, \ldots, u_k\}$ located in the same cell and interested in retrieving a set of chunks $C = \{c_1, c_2, \ldots, c_n\}$ from a file F, representing mobile cloud application content, which consists of N chunks. Geographic and location based data can be accessed with different levels of details [29], which in turn defines the size of the file F. For example, geographic maps are very large objects. They are usually split into small tiles, also called chunks [30], and sent to the client only for the selected zoom level, to avoid transmission of the whole map at once.

Each user requests a set of chunks $C_k \in C$ of the length n_k. The intersection of $C_1 \cap C_2 \cap \cdots \cap C_k$ defines a set of the length n of common chunks requested by all the users. As a result, the following relations hold for n:

$$\begin{cases} 0 < n \leq N, \\ n_k \geq n. \end{cases} \quad (1)$$

According to (1) it is possible to consider (a) the case in which all users are interested in the same content ($n = N$) and (b)
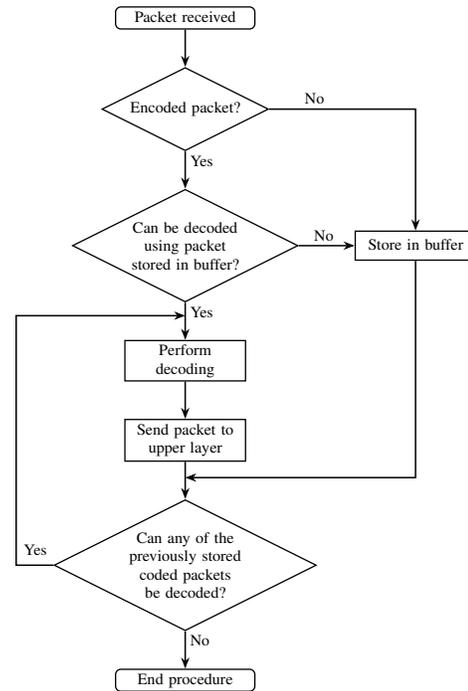

Fig. 5. Procedure executed at UE side after packet arrival.

the case in which the users are interested in parts of the same content ($n < N$), for instance in just a part of a map update.

In our model, mobile users send requests for content to the cloud. Each content request contains the list of chunks each user wants to retrieve. After the demand for chunks becomes known, eNodeB determines the set of n common for all the users chunks and computes the best allocation of coding opportunities. Having all content requests known at the beginning allows the proposed model to estimate the best gain of the applied network coding solution. However, in reality user requests arrive at different times requiring eNodeB to identify coding opportunities on the fly in real time. When possible, content chunks are encoded in pairs using XOR operation applied onto the chunk's content. The information on chunk IDs is then appended to IP packet header. Finally, coded packets are broadcast to all the cell users.

To deliver the set of n common chunks to k users, eNodeB performs network coding and transmits σ chunks:

$$\sigma = \begin{cases} \frac{n}{k} \cdot (k + \vartheta), & \text{if } r = 0; \\ \left\lfloor \frac{n}{k} \right\rfloor \cdot (k + \vartheta) + k + (r - 1), & \text{otherwise,} \end{cases} \quad (2)$$

where ϑ is the number of broadcast encoded transmissions and r is a remainder of the ratio n/k.

Fig. 6 shows a case of optimal allocation for content distribution. The i-th chunk destined to user j is denoted as $c_{i,j}$. The chunks destined to individual users are transmitted in unicast mode and denoted in light gray, while encoded chunks which are sent in broadcast are represented with dark gray. When $r = 0$, the transmission of k individual chunks is followed by the transmission of $\vartheta = k - 1$ encoded packets for n/k times. However, when $r \neq 0$ the remaining chunks are delivered with individual $k + (r - 1)$ transmissions.
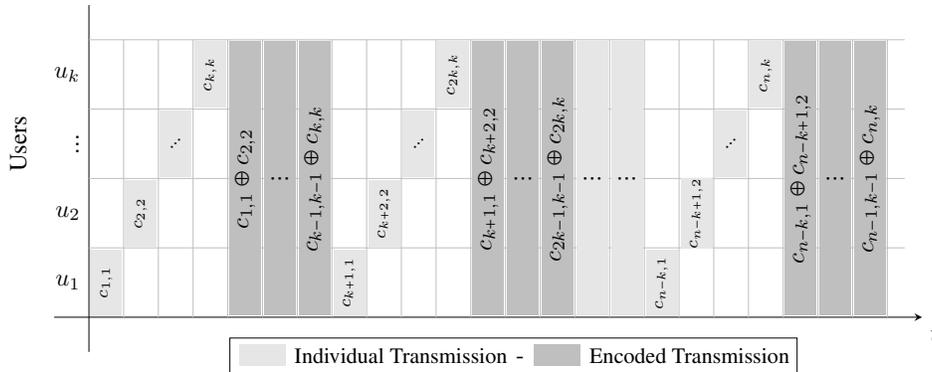
Fig. 6. Content distribution of n chunks to k interested users.

In case when network coding is not enabled, eNodeB transmits a total number of chunks equal to:

$$\gamma = k \cdot n. \qquad (3)$$

Analysis of *coding gain* helps to measure the benefit of using NC-CELL. Coding gain is obtained by computing a ratio between the total number of chunks transmitted without network coding and with NC-CELL enabled: $\eta = \gamma/\sigma$.

For example, with $n = 9$ and $k = 4$, according to (2) and (3), σ and γ are 18 and 36 respectively, which means the traditional approach with no network coding will require twice the number of transmissions ($\eta = 2$) with NC-CELL enabled.

Summarizing, NC-CELL approach is confirmed to be able to reduce the number of transmitted chunks in the cell considerably by using network coding techniques to combine chunks destined to different users.

IV. PERFORMANCE EVALUATION

For performance evaluation, we consider a system with 10 mobile users (k), each interested in downloading of up to 1000 chunks (n) of data. The small number of users is taken to underline the effectiveness of the proposed NC-CELL solution. Indeed, in real systems the number of users served by the same eNodeB that are interested in the same content can be limited. The main performance metrics are the number of transmissions performed by eNodeBs, which determines throughput of the system, and coding gain, which measures direct benefit from combining chunks together.

Fig. 7 shows the number of transmitted chunks for increasing n and k. It becomes obvious that the use of network coding (NC-CELL) can significantly reduce the number of transmissions performed by eNodeB over radio channel to deliver the same amount of content if compared with the NC-CELL disabled case. For example, for 10 users retrieving 100 common chunks, the eNodeB performs 181 transmissions with NC-CELL, which is comparable to the case with only 2 users when network coding is not used.

To quantify the benefits provided by NC-CELL, Fig. 8 presents coding gains as a function of n for variable number of users k. For $n \gg k$, the coding gain is proportional to k, as NC-CELL achieves good performance almost independently of the number of common chunks, but rather based on the number
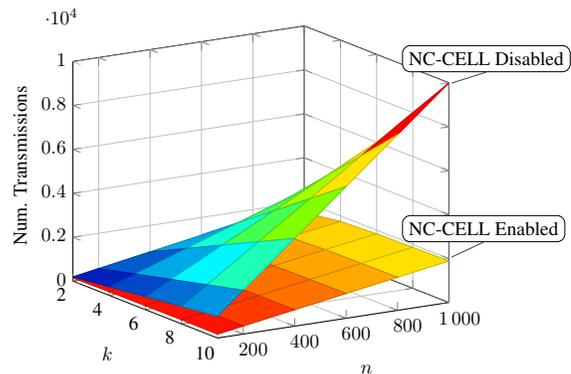

Fig. 7. Number of transmissions from eNodeB.

of users served by the eNodeB. For n values comparable to k, coding gain oscillations can be observed which amplitude grows proportionally to k. These oscillations are caused by r, the remainder of the ratio n/k, to the coding gain. For example, for $n = k = 10$ and no network coding, eNodeB transmits 100 chunks. On the contrary, with NC-CELL, eNodeB transmits only 19 packets: 10 individual packets are followed by the transmission of 9 encoded packets. This leads to $\eta = 5.26$. Now consider the case with $n = 11$ and $k = 10$. The users requested only one additional chunk with respect to the previous scenario. However, it will require 110 chunks to be transmitted for NC-CELL disabled case and only 29 chunks with NC-CELL, which corresponds to the coding gain of $\eta = 3.79$. The amplitude of these oscillations reduces with an increase of n, and approaches the straight line for $n \gg k$. This means that with a fixed number of users k, the coding gain becomes bounded in terms of n. On the contrary, having fixed the number of chunks n, the coding gain scales with the number of users.

To summarize, the performance of NC-CELL depends only on the number of users k and on the number of common chunks they want to retrieve n. Also k has higher influence on the scale of the gain with respect to n.

V. CONCLUSION

In this paper we proposed NC-CELL, a technique for efficient content distribution for cloud applications in mobile cellular
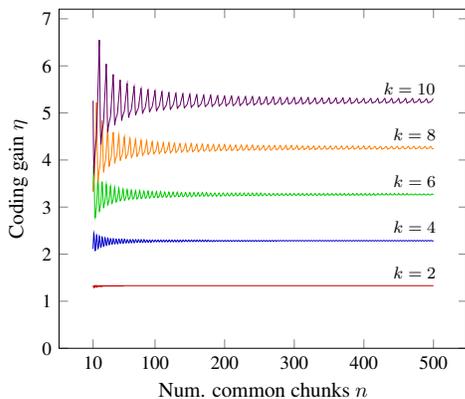
Fig. 8. Coding gain η.

networks. NC-CELL is based on network coding, a technology allowing to combine multiple packets into a single broadcast transmission which is sent over the radio interface. The proposal enables eNodeB nodes to monitor in transit traffic and cache it. Upon detection of coding opportunities, eNodeBs combine packets and deliver them in broadcast to the users.

The performance of the technique has been validated through analysis of the number of transmissions performed by eNodeBs and of the coding gain. The NC-CELL approach provides most of the benefits in systems with a large number of users and a large size of the common content that they are interested to receive.

As a future work, we will extend NC-CELL to distributed scenario. For this purpose, NC-CELL will be adapted to exploit LTE-A radio relay nodes. Both eNodeBs and radio relay nodes will use RLNC and not XOR coding to deliver mobile cloud applications content.

ACKNOWLEDGMENT

REFERENCES

[1] "Global mobile statistics 2013," http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats, 2013.

[2] "Cloud applications are more popular than you may think," http://callcenterinfo.tmcnet.com/Analysis/articles/358041-cloud-applications-more-popular-than-may-think.htm, October 25, 2013.

[3] "Mobile cloud computing industry outlook report: 2011-2016," http://goo.gl/qD51Fr, 2011.

[4] "Google Now," http://www.google.com/landing/now/, 2013.

[5] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: practical wireless network coding," in *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 4, 2006, pp. 243–254.

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

[7] T. Matsuda, T. Noguchi, and T. Takine, "Survey of network coding and its applications." *IEICE Transactions*, vol. 94-B, no. 3, pp. 698–717, 2011.

[8] J. Li, B. Jia, Z. Hu, and Y. Wang, "Design of joint network and channel coded relay in LTE-A system," in *2011 IEEE 13th International Conference on Communication Technology (ICCT)*, Sept 2011, pp. 1124–1129.

[9] Z. Lv, K. Xu, and Y. Xu, "A practical HARQ scheme with network coding for LTE-A broadcasting system," in *2012 International Conference on Wireless Communications Signal Processing (WCSP)*, 2012, pp. 1–6.

[10] Y. Lang, D. Wubben, A. Dekorsy, V. Braun, and U. Doetsch, "Improved HARQ based on network coding and its application in LTE," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, 2012, pp. 1958–1963.

[11] D. Vukobratovic, C. Khirallah, V. Stankovic, and J. Thompson, "Random network coding for multimedia delivery services in LTE/LTE-Advanced," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 277–282, 2014.

[12] S.-Y. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[13] A. Tassi, G. Rigazzi, C. Khirallah, D. Vukobratovic, F. Chiti, J. Thompson, and R. Fantacci, "Reliable multicast in LTE-A: an energy efficient cross-layer application of network coding," in *2013 20th International Conference on Telecommunications (ICT)*. IEEE, 2013, pp. 1–5.

[14] P. Vingelmann, M. V. Pedersen, F. H. Fitzek, and J. Heide, "On-the-fly packet error recovery in a cooperative cluster of mobile devices," in *2011 IEEE Global Telecommunications Conference (GLOBECOM 2011)*, 2011, pp. 1–6.

[15] H. Hamdoun, P. Loskot, T. O'Farrell, and J. He, "Practical network coding for two way relay channels in LTE networks," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5.

[16] Y. Hu, H. C. Chen, P. P. Lee, and Y. Tang, "Nccloud: applying network coding for the storage repair in a cloud-of-clouds," in *USENIX FAST*, 2012.

[17] D. S. Papailiopoulos, J. Luo, A. G. Dimakis, C. Huang, and J. Li, "Simple regenerating codes: network coding for cloud storage," in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 2801–2805.

[18] X. Jin and Y.-K. Kwok, "Cloud assisted P2P media streaming for bandwidth constrained mobile subscribers," in *2010 IEEE 16th International Conference on Parallel and Distributed Systems (ICPADS)*, 2010, pp. 800–805.

[19] L. Keller, A. Le, B. Cici, H. Seferoglu, C. Fragouli, and A. Markopoulou, "Microcast: cooperative video streaming on smartphones," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 2012, pp. 57–70.

[20] D. Zeng, S. Guo, I. Stojmenovic, and S. Yu, "Stochastic modeling and analysis of opportunistic computing in intermittent mobile cloud," in *2013 8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2013, pp. 1902–1907.

[21] "3GPP - LTE," http://www.3gpp.org/LTE/.

[22] S. N. Shahab, T. S. Kiong, and A. A. Abdulkafi, "A framework for energy efficiency evaluation of LTE network in urban, suburban and rural areas," *Australian Journal of Basic & Applied Sciences*, vol. 7, no. 7, pp. 404–413, 2013.

[23] R. K. Jain, S. Katiyar, and N. K. Agrawal, "Hierarchical cellular structures in high-capacity cellular communication systems," *CoRR*, vol. abs/1110.2627, 2011.

[24] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences," *Proceedings of the National Academy of Sciences*, vol. 107, no. 52, pp. 22436–22441, 2010.

[25] "myCloudLOCAL," http://www.mycloudlocal.co.uk/, 2013.

[26] P. Vingelmann, M. V. Pedersen, F. H. Fitzek, and J. Heide, "Multimedia distribution using network coding on the iphone platform," in *Proceedings of the 2010 ACM multimedia workshop on Mobile cloud media computing*. ACM, 2010, pp. 3–6.

[27] M. Pedersen and F. Fitzek, "Implementation and performance evaluation of network coding for cooperative mobile devices," in *ICC Workshops '08. IEEE International Conference on Communications Workshops, 2008.*, May 2008, pp. 91–96.

[28] B. Gajic, J. Riihijärvi, and P. Mähönen, "Performance evaluation of network coding: effects of topology and network traffic for linear and XOR coding," *Journal of Communications*, vol. 4, no. 11, pp. 885–893, 2009.

[29] A. Çöltekin and T. Reichenbacher, "High quality geographic services and bandwidth limitations." *Future Internet*, vol. 3, no. 4, pp. 379–396, 2011.

[30] M. Adnan, A. Singleton, and P. Longley, *Developing efficient web-based GIS applications*. Centre for Advanced Spatial Analysis (UCL), 2010.