2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies

Streaming at the edge: Local service concepts utilizing Mobile Edge Computing

Olli Mäkinen
Department of Computer Science and Engineering
Aalto University School of Science
Espoo, Finland
Email: olli.makinen@aalto.fi

***Abstract*—Mobile Edge Computing is a novel concept where regular servers is placed within the LTE networks' eNode-Bs. These servers allow to run third party services directly at the edge of the network, providing minimal latency to the user equipment while minimizing the use of packet core bandwidth. Conserving bandwidth has become increasingly important for the mobile network operators as the mobile traffic volumes have skyrocketed, and forecasts suggest that the volume growth will continue in the future as well. In this study, novel local streaming service concepts based on Mobile Edge Computing and their business models are proposed through explorative research, utilizing a relevant theoretical framework. Furthermore, future research for the concepts are suggested.**

***Index Terms*—Mobile Edge Computing, mobile services, mobile service business models**

I. Introduction

The amount of data transferred in mobile networks has been rising at a very high rate during the recent years, with no end in sight. According to Cisco, the mobile data traffic grew in 2014 by 69 percent, and video traffic amounted to 55 percent of the total. It is also expected to grow at an average of 57 percent every year until 2019. [1]

This kind of a growth has created and will continue to create increasing pressure for the Mobile Network Operators (MNOs) to invest heavily in infrastructure and to find ways to limit the bandwidth use of the third-party OTT (Over-the-Top) services, such as video streaming and Peer-to-Peer traffic. These ways are often limited by regulation, which is rather heterogenous throughout the world. [2]

A recent industry initiative coordinated by ETSI, Mobile Edge Computing (MEC), aims to create possibilities for bandwidth saving through the operators' packet cores as well as to reduce network-induced delay. The MEC concept builds upon physical servers located at the LTE networks' eNode-Bs (eNB), equivalent to base stations in legacy networks. These relatively generic servers allow third party services to run at the very edge of the mobile network, providing computational and storage capacity for a variety of different services. It also aims to create a standardized ecosystem to facilitate cooperation between the MNOs and OTT service providers. [3]

In this study, two service concepts and their business models are proposed which build upon MEC, utilizing a common framework. Furthermore, as the MEC initiative is at a very early stage, the concepts and their success factors are analyzed, creating a basis for further research.

II. Background literature

As Mobile Edge Computing is at the moment quite industry-driven and new, the existing research is extremely sparse. One of the first studies involved computational offloading for VoLTE utilizing MEC servers[4]. In this study the power consumption of an user equipment (UE) was successfully decreased when using software encoding.

A. Mobile cloud computing

Mobile Edge Computing shares some aspects with the more researched Mobile Cloud Computing (MCC) concept. However, one significant difference exists; in MCC, usually the computing power is located somewhere in the Internet, producing significant transmission latency which can be a hindrance for a multitude of services, for example gaming[5].

Deploying cloudlets, or smaller but dispersed versions of full clouds, produces another concept called fog computing, optimally resulting in less transmission latency and more robustness for mobile offloading[6]. Mobile fog computing can also function through WLAN connections, which simplifies demands for the mobile network[7].

MEC could be seen as the ultimate version of mobile fog computing, as the computational power is as close as possible to the UE without utilizing WLAN.

B. M2M research

Machine-to-machine (M2M) is also currently a hot topic in the industry. Usually M2M describes a concept where sensors communicate to each other and centralized systems, but the research is relevant for MEC as well. Mobile fog has also been described as having a focus for sensor networks[8].

Latency in M2M communication has been studied comprehensively by Nikaein and Krco[9], where it is noted that the enhanced packet core (EPC) and IP backbone can contribute a

978-1-4799-8660-6/15 $31.00 © 2015 IEEE
DOI 10.1109/NGMAST.2015.35

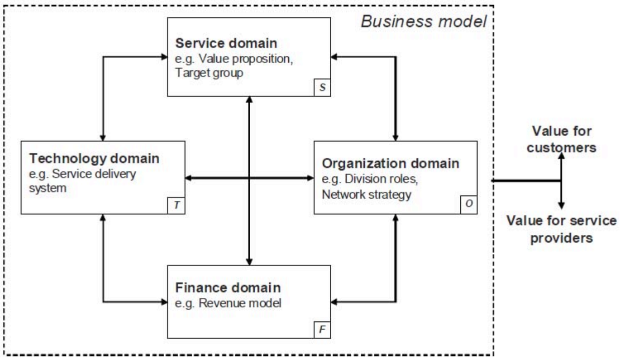

Fig. 1. STOF business model domains[13]

significant latency - this emphasizes the importance of MEC. HSPA and LTE latencies have also been compared[10]. In this study, it was noted that the roundtrip average latency in LTE networks was lower, but uplink latencies were normally higher than with HSPA.

C. ALTO

ALTO, standing for Application Layer Traffic Optimization, is an IETF working group aiming to produce a protocol for optimizing peer-to-peer traffic[11]. In ALTO, a server provided by the ISP gives clients directions to the nearest location of a requested resource. These directions can take for example pricing as well as the network topology into account.

ALTO has not yet been researched very much in the mobile domain. In one publication, ALTO is enhanced for mobile use, taking the mobility circumstances into account[12].

III. Research Methodology

Given the very early nature of Mobile Edge Computing, native applications are difficult to implement and measure. Practical real-world applications can still take some time before the infrastructure development and standardization processes are mature enough. This study was conducted as an explorative case study research, using brainstorming within a small group familiar with the MEC concept to innovate key differentiation features which are enabled through MEC.

The STOF model by Bouwman et al. was used as the theoretical framework for describing the novel mobile service business models. STOF, standing for Service, Technology, Organization and Finance, emphasizes the holistic of a business model, with four seemingly separate domains connected to each other. This model is especially useful at the experimentation stages of service design, as it considers also the delivery process and financial viability instead of just concentrating on the user experience. [13]

The four domains and their interconnections are shown in Figure 1.

A. STOF: Service

The service domain consists of issues related to the intended vs. perceived user value of the service with a focus on the mobile aspects: ubiquity, personalization, context-awareness and interaction. End-user pricing is also an important aspect related to the perceived and expected user value. Additionally, privacy issues have become more and more important lately. [13]

B. STOF: Technology

Requirements derived from the service domain specify the technical architecture of the service. Aspects regarding the user devices, network and data infrastructure as well as the interconnection to other required services are included in the technology domain. [13]

C. STOF: Organization

In the organization domain, the complete value network of the service is described. In a mobile service, this network of actors might be very complex with each of them having their own strategies and goals. Additionally, the interaction and value creation direction between actors need to be clearly defined. [13]

D. STOF: Finance

The finance domain describes the financial arrangements between the actors. Much more than just basic user-paid revenue models, the issues include investments, risk management and cost reduction as well. The ultimate goal in the finance domain is to create a balanced design which is profitable for everyone in the value network. [13]

IV. Proposed Service Concepts

According to Cisco, it is estimated that the share of mobile video streaming of the total mobile data traffic will increase even higher than of today in the future[1]. Thus, technology that helps reduce the video stream traffic through the packet core is of substantial interest. Both of these proposed service concepts are built upon local video streaming.

A. P2P communication

This service concept is somewhat similar to the existing study with VoLTE[4], but is based on a different technology and does not involve computational offloading. In this concept, peer-to-peer data traffic is routed by the eNB and MEC server to either directly to another UE within the same eNB or via the X2 links to a UE within an adjacent eNB. The X2 links, normally used for control data, connect nearby eNBs together[14].

This allows OTT services to allow direct communication between end users without going through Internet, thus saving the network operator's backbone capacity and traffic. Additionally, this saves the OTT's own backbone capacity and traffic as well, creating a viable business case.

In this concept, video streaming or calling can be considered to be the prime use case, but it can be extended to other areas as well, such as mobile gaming. Connecting the MNOs and OTT service providers together has been researched where the MNO provides TURN for the OTT services[15]. A TURN service located within the MEC server could be used by WebRTC video P2P communications.

B. Event video streaming

In this concept, a local breakout functionality from the MEC server is used to digest video streams from a localized video production studio. The video streams are then transferred directly via the LTE network to the UEs, enabling users to watch the stream in real-time. This would be especially useful in event and sports venues where visibility to the centre of action is limited. Multiple video angles provide a source of information previously not available to the spectators at all – usually venues provide one single video wall or a media cube which contains content selected by the director.

This has already been trialed in practice[16], but this concept also involves interaction between the spectators and organizers as well as user-created streams from UEs. Spectator interaction has been noted to act as a co-creator of the whole event atmosphere[17].

V. Business model analysis

A. Service design

1) P2P communication: In services aimed for local use, such as messaging services Badoo and Tinder as well as local gaming, users communicate with other users who are geographically close by. In other cases, such as group photo messaging as well as video and voice calling, the target group of users can also reside either within the same eNB or in a close radius.

In this concept, as the OTT service finds out the target users based on for example the phone number, it relays these to a P2P API located within the eNB, which then directs the data traffic intelligently through the shortest path. Control traffic is relayed to the OTT via EPC and Internet.

Mobile multiplayer games benefit from the decreased latency as data traffic can go directly from one peer to another, skipping EPC and Internet altogether. MEC also allows mobile offloading for gaming, as there can be a cloudlet gaming service running in the MEC server. The latency in LTE is small enough to justify this[18], especially when compared to HSPA[10].

The location-specific OTT services could also benefit from the location data provided by the MEC server, as it natively knows relevant UE signal strength data. As the phone's own GPS is not required as much anymore, battery savings are imminent benefiting the end user.

The main value for the operator as well as the OTT provider is cost savings. The end user experiences lower latency especially in video streaming and gaming services as well as longer battery life through the MEC provided location data and possible offloading for better quality graphics[18].

2) Event video streaming: In this concept, the end user experiences real-time video streams from the event he or she is participating in. As the video streams are transferred via low-latency LTE, the user does not experience similar delays as with Internet-relayed video streams.

The video production studio, working in a close relationship with the MNO, generates multiple video streams from different angles and feeds them directly into the eNB. This allows the audience to select their desired camera feed, which can be focused on the action or perhaps even somewhere else which normally wouldn't be shown on TV constantly (e.g., coaches). The application can also show detailed player biographies or game statistics, which would be cumbersome to follow in another way.

Interactivity between the spectators and the event can also be provided. Voting of played music or desired camera angles can be feed back to the DJ and video production, providing a deeper experience than a simple and passive one-way feed[17]. Additionally, the UEs can also upload camera streams in a similar manner to provide even more event co-creation. However, user-provided streams need some control to disallow inappropriate content.

The problem of having only one MNO providing this service is solved by using device rentals at the event. Suitable and rugged enough LTE tablets or smartphones are rented to the end user for the duration of the event. This also simplifies the usage of the application, as the devices are pre- configured and locked. This application also can show advertisements, as are typical for sports events. On the other hand, being the only operator that provides this service also acts as a differentiation bonus compared to the competitors.

B. Technology design

1) P2P communication: This concept requires somewhat elaborate services running within the MEC server(s), such as traffic analysing and intelligent forwarding. An API is provided to the OTT providers, which automates the traffic flow, going through direct X2 links if they are the fastest option, alternatively through EPC if more direct X2 links are not available when the last option is through the Internet if neither of these is viable.

The traffic flow is automated for example by using the phone number as the UID. The UID can be injected into the TCP headers, which are interpreted by the traffic analyser. This functionality is preferably completely hidden from the developer. In the case of an eNB handover, the X2 links retain and cache the data in a sufficient matter. It is also possible to utilize ALTO (Section II-C) as an already standardized approach to find peers. ALTO could use the phone number or user name as the resource, and the MNO would provide an ALTO server within their network. ALTO has been studied as having a positive effect on local P2P video connections[19].

Lawful interception (LI), i.e., the legal requirement to forward information in a telecommunications network to the authorities, is enabled through local caching of the legally required data and uploading to the central handover interface during off-peak hours. The requirements for lawful interception vary regionally. If required, certain subscribers can be monitored continuously based on a list. If legislation does not allow for caching of the extracted information and still requires everything to be stored in real-time, the benefits of this concept to the MNO are very far from optimal.

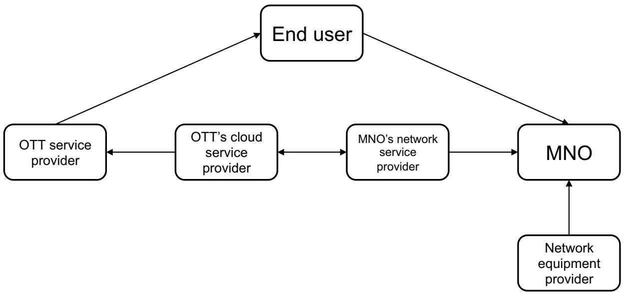

Fig. 2. An example of a P2P communication service value network

2) Event video streaming: The technology domain relies heavily on the local video production studio. At the MEC server, a suitable application is running, relaying the video streams into the LTE network for UEs as well as possibly receiving the video streams from UEs. The only outside link is to the advertisers, who can showcase selected material on the applications for example during pauses. The advertisement material, e.g., video clips, can be stored inside the MEC server to further optimize the traffic usage.

The video production unit will provide encoded video streams for the relaying software inside the MEC server, which will then broadcast them through the eNB utilizing LTE's MBMS mode[20], [21]. A hybrid cellular and ad-hoc method has also been proposed[22].

The user application provides the stream selection functions as well as the rich data on the players or statistics of the sports event. Interactive features can also be added, which will then feed back to the video production and event hosting team. Advertisements can be broadcast in a similar fashion as the actual video streams to the user applications, forcing the screen to show them during pauses if deemed necessary by the organizers.

C. Organization design

1) P2P communication: The organization domain consists of the end user, MNO and OTT service providers as well as the network service providers, equipment providers and the Lawful Interception agency (see Figure 2). The central actor in this case is the OTT service provider, who initially proposes the agreement to the MNO. Ideally, this MEC API would be standardized so that it is relatively easy for OTT providers to start using it. If required, the LI agency is also part of the organization, where legally required data is transferred to the handover interface.

2) Event video streaming: The organization design is rather complex. The central actor is the event organizer, who creates the agreements with the other actors – the MNO, end users and the video production team as well as the advertisers/sponsors. In the case of device rentals, the rental agency is very much involved in the process as well. These kind of events are most probably already being recorded and streamed outside of the venue already, so the video production studio only needs to be included in the video orchestration (Figure 3).

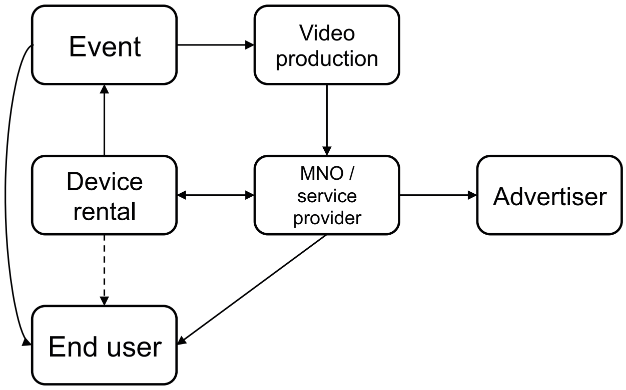

Fig. 3. An example of an event video streaming service value network

The value flow can be very easily attributed to the end user. They benefit from this kind of a service, are accustomed to advertisements due to the nature of sports events and do not necessarily consider the device rental costs to be of an issue. This also generates a lot of value to the MNO, as they can promote and distinguish themselves very well from the competition. Additionally, sports events usually only have one MNO as a sponsor, providing even more differentiation.

D. Finance design

1) P2P communication: In the financial domain, the main issue is whether the cost savings are enough for the parties to invest into the technology. However, benefits are clear for most of the parties. The end user benefits from less latency and battery consumption and the OTT service provider as well as the MNO from smaller traffic costs. Caching of the LI data can also balance out the MNO's network usage.

The OTT service provider pays a monthly fee for using the service, which would be less than their cost savings in infrastructure and traffic costs. It would be ideal to have multiple different OTT service providers using the same technology to create revenue for the MNO. The MNO also invests into the technology and licensing from the network equipment provider to save their traffic costs paid to the MNO's network service provider.

2) Event video streaming: In the finance domain, it is clear that all of the parties gain enough revenue to support investments. In such events, rugged enough LTE tablets or smartphones are financially viable investments due to the value provided by the short-term rental to the end user. As the prices are going lower all the time, even their purchase price to the device rental agency is not going to be a problem. However, this requires the events to happen rather often (e.g., an ice hockey stadium), and might not work with venues with limited use throughout the year (e.g., a Formula 1 track).

The end user will provide revenue from the device rentals or a single/monthly fee to their MNO as well as from the event tickets. This service could also be included as a perk in for example season tickets. The advertisers can provide interesting content or even competitions directly to the end user, paying the MNO or the event organizer for premium content delivery.

VI. Drivers and restraints

In order for these concepts as well as the MEC in general to succeed, there are drivers which are essential as well as restraints which can hinder the progress.

A. Drivers

One key driver is the **standardization** process within MEC. Common APIs among the network equipment manufacturers, as well as standardized virtual server specifications are required for an ecosystem to be built up. Additionally, **trusted brokers** between the MNOs and OTT service providers need to exist, as otherwise agreements between all of the players in the ecosystem will be difficult to achieve. As the services run very closely to the mobile network itself, it is imperative to have the **security** issues ironed out. However, this means that updating the service is a tedious process, where each iteration will need to be checked according to the local telecom security regulations. Additionally, as the concept is entirely new, finding **early adopters** is critical to achieve sufficient interest in the global mobile services industry. To facilitate early adoption, there should be good enough **incentives** for both the service providers as well as the MNOs to facilitate cooperation.

B. Restraints

One of the most important possible restraints in the adoption of MEC services is the **slow deployment of server hardware** to the eNode-Bs. This is a classic chicken-and-egg problem, where it is not viable for the MNOs to invest in MEC hardware unless there are services which are economically viable and vice versa. One solution would be to have generic enough services running on the platform which are beneficial for the operator, e.g., analysis tools for that particular eNode-B. **Pricing** can also one key restraint which will affect the concept adoption for the OTT service providers in addition to the MNOs. Additionally, if regulation requires all of the traffic in a mobile network to be forwarded to a certain location, such as in the case of **lawful interception**, it can make the adoption less attractive to the MNOs as the reduction of traffic within the packet core is effectively non-existent. If the lawful interception traffic does not need to be real-time, it could be stored within the MEC units and transferred at off-peak times. The lawful interception problem was not addressed in the as of yet only published study[4].

VII. Conclusion and future work

As the mobile network operators struggle to cope with increased data traffic rates, it is crucial to find ways to limit the growth without affecting the level of user experience. Mobile edge computing shows potential not only in decreasing the data traffic amounts in the EPC and IP backbone, but additionally decreasing latency and thus enhancing the user experience.

In this study, two new MEC-based service concepts and their business models are presented by using a framework designed for mobile services. Based on them, drivers and restraints for the success of these service concepts are described.

However, as MEC is taking its first steps, testing the practicality and performance gain of these concepts has not yet been possible. These are obvious targets for future research, as well as the complete value network of MEC enabled services including service brokers. ALTO in the mobile environment as well as in the MEC context should also be researched. As the MEC standardization process continues, it can be expected that the interest in MEC services in the research community will increase further.

References

[1] Cisco Systems. (2015, February) Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014-2019. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf

[2] K. Jayakar and E.-A. Park, "Emerging frameworks for regulation of over-the-top services on mobile networks: An international comparison," in *Proceedings of the 42nd Research Conference on Communication, Information and Internet Policy (TPRC)*, 2014.

[3] ETSI. (2014, September) Mobile-edge computing - introductory technical white paper. [Online]. Available: https://portal.etsi.org/Portals/0/TBpages/MEC/Docs/Mobile-edge_Computing_-_Introductory_Technical_White_Paper_V1%2018-09-14.pdf

[4] M. Beck, S. Feld, A. Fichtner, C. Linnhoff-Popien, and T. Schimper, "ME-VoLTE: Network functions for energy-efficient video transcoding at the mobile edge," in *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on*, Feb 2015, pp. 38–44.

[5] U. Lampe, R. Hans, and R. Steinmetz, "Will mobile cloud gaming work? findings on latency, energy, and cost," in *Proceedings of the 2nd International Conference on Mobile Services (MS 2013)*, June 2013, pp. 960–961.

[6] M. Satyanarayanan, R. Schuster, M. Ebling, G. Fettweis, H. Flinck, K. Joshi, and K. Sabnani, "An open ecosystem for mobile-cloud convergence," *Communications Magazine, IEEE*, vol. 53, no. 3, pp. 63–70, March 2015.

[7] T. H. Luan, L. Gao, Z. Li, Y. Xiang, and L. Sun, "Fog computing: Focusing on mobile users at the edge," *CoRR*, vol. abs/1502.01815, 2015. [Online]. Available: http://arxiv.org/abs/1502.01815

[8] K. Hong, D. Lillethun, U. Ramachandran, B. Ottenwälder, and B. Koldehofe, "Mobile fog: A programming model for large-scale applications on the internet of things," in *Proceedings of the Second ACM SIGCOMM Workshop on Mobile Cloud Computing*, ser. MCC '13. New York, NY, USA: ACM, 2013, pp. 15–20. [Online]. Available: http://doi.acm.org/10.1145/2491266.2491270

[9] N. Nikaein and S. Krea, "Latency for real-time machine-to-machine communication in LTE-based system architecture," in *Wireless Conference 2011 - Sustainable Wireless Technologies (European Wireless), 11th European*, April 2011, pp. 1–6.

[10] M. Laner, P. Svoboda, P. Romirer-Maierhofer, N. Nikaein, F. Ricciato, and M. Rupp, "A comparison between one-way delays in operating HSPA and LTE networks," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012 10th International Symposium on*, May 2012, pp. 286–292.

[11] J. Seedorf, S. Kiesel, and M. Stiemerling, "Traffic localization for p2p-applications: The alto approach," in *Peer-to-Peer Computing, 2009. P2P '09. IEEE Ninth International Conference on*, Sept 2009, pp. 171–177.

[12] T. Taleb, E. Ndih, and S. Cherkaoui, "On supporting mobile peer to mobile peer communications," in *Communications (ICC), 2012 IEEE International Conference on*, June 2012, pp. 596–600.

[13] H. Bouwman, H. de Vos, and T. Haaker, *Mobile Service Innovation and Business Models*, 1st ed. Springer Publishing Company, Incorporated, 2010.

[14] A. Larmo, M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner, and H. Wiemann, "The LTE link-layer design," *Communications Magazine, IEEE*, vol. 47, no. 4, pp. 52–59, April 2009.

[15] S. Becot, E. Bertin, J.-M. Crom, V. Frey, and S. Tuffin, "Communication services in the web era: How can telco join the OTT hangout?" in *Intelligence in Next Generation Networks (ICIN), 2015 18th International Conference on*, Feb 2015, pp. 208–215.

[16] Optus. (2015, March) Optus & Nokia Networks pool expertise to trial real-time video and multimedia content delivery through Liquid Applications. [Online]. Available: https://media.optus.com.au/media-releases/2015/optus-nokia-networks-pool-expertise-to-trial-real-time-video-and-multimedia-content-delivery-through-liquid-applications/

[17] M. Ludvigsen and R. Veerasawmy, "Designing technology for active spectator experiences at sporting events," in *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10)*, 2010, pp. 96–103.

[18] T. Kamarainen, M. Siekkinen, Y. Xiao, and A. Yla-Jaaski, "Towards pervasive and mobile gaming with distributed cloud infrastructure," in *Network and Systems Support for Games (NetGames), 2014 13th Annual Workshop on*, Dec 2014, pp. 1–6.

[19] K. Teket and M. Sayit, "P2p video streaming with alto protocol: A simulation study," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2013 IEEE International Symposium on*, June 2013, pp. 1–6.

[20] O. Oyman, J. Foerster, Y. joo Tcha, and S.-C. Lee, "Toward enhanced mobile video services over WiMAX and LTE [WiMAX/LTE Update]," *Communications Magazine, IEEE*, vol. 48, no. 8, pp. 68–76, August 2010.

[21] C. Khirallah, D. Vukobratovic, and J. Thompson, "Bandwidth and energy efficiency of video broadcasting services over LTE/LTE-A," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, April 2013, pp. 1909–1914.

[22] N. M. Do, C.-H. Hsu, J. Singh, and N. Venkatasubramanian, "Massive live video distribution using hybrid cellular and ad hoc networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*, June 2011, pp. 1–9.