Fuelling Big Data Intelligence into Future Multimedia System: Reflection and Outlook

Yonggang Wen
School of Computer Engineering
Nanyang Technological University
Singapore
ygwen@ntu.edu.sg

Wenwu Zhu
Department of Computer Science
Tsinghua University
Beijing, China
wwzhu@tsinghua.edu.cn

***Abstract*—The last decade has witnessed an explosive growth in multimedia big data, with the invention of big data analytics technologies, such as Hadoop, Spark, Storm. All these technologies are valuable to mine intelligence from multimedia big data, and further shed new insights into multimedia networks. This paper presents an outlook on the development of future multimedia networks and discuss utilizing the big data intelligence into the design and optimization. We start with an end-to-end view for future multimedia networks, including contents, multimedia networks, and consumers. For each part, we enumerate several representative applications and technologies that bring new challenges and opportunities. After that, we discuss the common idea of utilizing the big data intelligence to benefit the whole ecosystem, and highlight some open research issues for the future.**

***Index Terms*—Big Data Analytics, Future Multimedia Communication, SDN**

I. Introduction

The era of big data [1] is upon on us, especially multimedia big data, bringing with it an urgent need for advanced multimedia networks and mechanisms. According to a report from Cisco, the IP video traffic will be 79 percent of all consumer Internet traffic in 2018, up from 66 percent in 2013. However, user experience is severely constrained by the rigescent network design philosophy. First, to cope with the continuously growth of user demand on rich media experience, more and more network equipments are being deployed. This effort can alleviate the peak pressure of the network infrastructure, but result in lower utilization ratio. Second, the relatively static mechanism of system resource provision in existing infrastructures cannot react fast enough to flash-crowd demands and various emerging multimedia applications. For example, videos propagated on social networks have distinctive dissemination behavior in comparison with the traditional VoD service. This calls for the new design for resolving the tussle between the growing demand of multimedia applications and the aforementioned limitations of existing multimedia networking infrastructures.

The design principle of multimedia networks follows a general rule [2]: according to user requirements and application characteristics, utilizing various enabling technologies to lower the system maintenance cost. There is a fundamental tradeoff between the system cost and user requirements. On one hand, multimedia networks should meet some basic features, including scalability, heterogeneity, reliability, usability, and security, and further be operated with an acceptable expenditure. Such cost will be imposed to end users and affect the penetration of multimedia service. On the other hand, users demand for an enhanced QoS and richer services, while paying for less. This tradeoff can be solved by emerging technologies and system optimization mechanisms.

Recently, the development of re-programmable technologies offers advanced solutions to reduce the cost of deploying and operating multimedia networks. In the computing and storage level, built on hardware virtualization and on-demand service provision paradigm, cloud computing is becoming a prevalent solution for industry service deployment. Under this paradigm, system resources can be instantiated into virtual machines, whose capacity can be dynamically tailored for customized media applications (e.g., transcoding, rendering, etc.). In the communication level, the emerging software defined networking (SDN) decouples the data plane and control plane, and adopts a centralized strategy to operate the overlay network. Under this paradigm, we can dynamically determine the optimal data forwarding path, and achieve a more efficient multimedia network.

Due to the open data policy, more and more Internet service providers (SPs), from CDN SPs (e.g., Akami) to video SPs (e.g., Youtube, Twitter), disclosure abundant application programming interfaces (APIs) for researchers. As a result, more and more system details can be known or inferred, which may transform the system optimization philosophy from the model based to learning based. In particular, the model based methods consist of two steps: 1) constructing an accurate mathematical model to describe the system running; 2) acquiring the solution by rigorous mathematical deriving. However, this type of methods incurs several drawbacks. For example, the system model may be in-accurate as the system is dynamic evolving. In contrast, using the knowledge mined from the open datasets or ecosystems, the learning based optimization methods can customize the system architecture accordingly, and devise optimization policies without accurate system models. For instance, by analyzing the records of video dissemination paths over social networks, we can find that videos are propagated in small cliques, in which social users are connected by social relationships, and have close

978-1-4799-8688-0/15 $31.00
DOI 10.1109/BigMM.2015.21

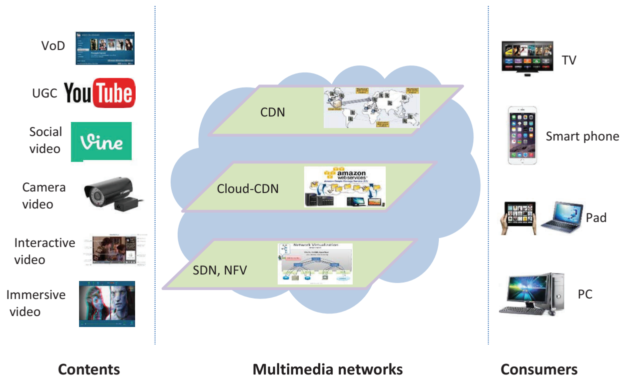

Fig. 1. An end-to-end view of multimedia communication networks, including contents, multimedia networks, and consumers.

geo-locations, as well as similar interests. We can build the community-based social video distribution and request routing policy [3] to optimize the system performance.

This paper discusses the development of future multimedia networks from aforementioned two aspects. We first describe the architecture of multimedia networks from an end-to-end perspective, including contents, networks, and consumers. For contents and networks parts, we presents representative applications which have different characteristics and require for specific solutions. For network part, we discuss several emerging technologies. After that, we highlight the concept of learning based system design and optimization.

II. Multimedia Communication Networks and Enabling Technologies

In Fig. 1, we illustrate a schematic end-to-end view [2] of the multimedia communication system. The system consists of three participatory stakeholders in the digital media value chain, including contents, media networks, and consumers.

- *Contents*: include all possible multimedia services and applications. In addition to the traditional VoD and UGC services, social video, camera video, interactive video and immersive video possess richer features and user experience, present overwhelming technical challenges for media systems, and demand new solutions.
- *Media Networks*: are responsible for media processing, analyzing, and transmitting. Currently, more than half of the video traffic is delivered by CDNs, built by CDN SPs and content providers. As the emerging of cloud computing, SDN and NFV, the underlaying CDN infrastructure is embracing these advanced technologies to reduce the operational cost with higher efficiency.
- *Consumers*: refer to the end devices that users are using for multimedia consumption.

For each part, we will discuss representative characteristics and technologies in the following sections.

A. Contents

Multimedia contents can be created by professional producers with sophisticated digital cameras or regular users who capture videos with their own devices. Considering the generation and distribution environments, different types of video services present different levels of challenges. In this work, we highlight four representative multimedia services:

Social Video refers to video propagated on social networks. For example, more and more users tend to watch videos on Youtube, and share them to Youtube or Twitter. In comparison with UGC videos, the unique feature of social videos comes from the information dissemination pattern on social networks, bringing along extra challenges to the social video distribution. In general, social videos spread along social relationships. As such, the social relationships and social activities determine the propagation of videos among users, rather than the recommendation in traditional UGC websites. Therefore, we can utilize the information propagation characteristics to assist video distribution [4]. For example, the social influence characteristics, e.g., follower number, interaction level, etc., can be exploited to predict the popularity of social videos; By clustering social users with social relationships and similar interests, we can infer whether a video will be watched by other social users, and further design video replication and cache algorithms.

Camera Video is generated continuously by the widely deployed network cameras for various purposes. The salient features include: 1) The volume of camera videos is huge. Taking an HD stream from a camera with bit rate 1.5 Mbps, the produced data over one day is 16.2 GB. Once using thousands or then thousands of cameras to monitor a vital area, such as Times Square, the stream data produced per day will stuff dozens of disks each with 4TB capacity; 2) The data formats of network cameras are quite different, including JPEG, MJPEG, MPEG-4, and the frame rates are also dissimilar. These features present various technical challenges in storing, processing and analyzing camera videos. However, the context information can be utilized to alleviate these difficulties. For instance, machine learning technology can be used to model the background and detect the object, and encode object and background in different data streams to increase the compression ratio; Another case is that video streams from adjacent cameras are correlated (e.g., the order of an object passing through an sequence of cameras can be inferred), one can use this relation to dynamically set priorities to different streams and achieve more efficient on-line analysis.

Interactive Video, also known as non-linear videos or multi-path videos, allows users to traverse different plot sequences, depending on their interactions with videos. These types of videos enable more personalized watching experience, but also present new challenges. Neither the content providers nor the user may not know in advance which paths a user will take. Therefore, the content replication strategies for service providers and video streaming strategies for clients should be revised accordingly. For service providers, as the popularity of different paths may be distinctive, they should, on one hand consider the popularity diversity of video chucks to strategically design replication policies to reduce storage cost, on the hand guarantee smooth playback experience. For playback clients, videos must be prefetched along multiple paths to ensure the playback is seamless [5], even the choice to the next path is made in the last possible moment. To achieve this goal, we need to re-design the streaming protocol between clients and video servers with full flexibility and less overhead, and the prefetching strategies to reduce the waste video chunks.

Immersive Video, also called stereoscopic 3D video [6], has become the most attractive multimedia contents after the shown of Avatar 3D movie. Although many major content producers have shipped their 3DTV products and several TV operators are launching 3D channels all around the world, there are still many challenges in 3D video representation, distribution, and user interaction. In general, 3D scene can be represented either in pure geometric data structure, or in a mixed format that is partially in pixel-based representation and partially in geometric-based description. Major approaches for 3D video coding can be categorized into fixed or free-viewpoint, e.g., two-view stereo video coding, multiple view coding, etc. The choice of coding schemes also depends upon the transmission system. Clearly, the 3D video needs more bandwidth to carry more views. In addition, the encoded 3D video includes more dependency due to the exploitation of inter-view prediction. The existing streaming mechanisms cannot be applied to 3D video distribution directly.

B. Multimedia Networks

Multimedia networks, mainly refers to content delivery networks, play a vital role in multimedia distribution. It is reported that more and more Internet video traffic will cross CDNs. To enhance the efficiency, the traditional CDNs are embracing advanced technologies.

CP-CDN: Although there are many mature commercial CDN solutions available, content providers (CPs) are building their own CDNs, named CP-CDN, to improve the quality of content delivery for a variety of services, including webpages, media streaming, social networks, etc. In the conventional CDN based solution, CPs are decoupled with the underlying CDN infrastructure. As a result, CPs have little control to the CDN to optimize the system performance, and CDN has little knowledge about the deployed service. In contrast, using the CP-CDN architecture, full knowledge about the underlying infrastructure and the upper layer services can be exploited to optimize the user experience and service quality. For webpage services, considering the fact that each webpage is dynamically synthesized with different components (e.g., html framework, css, multimedia objects, etc.), we can prioritize the download sequence for different components, and further strategically select peer servers with diverse service delay [7]. For social video services, we can collect the user background information, including followers, followees, social relationships and content access records, to design the social influence based popularity prediction algorithms, and proactively replicate contents, so that users can fetch the contents from nearby CDN servers [7].

Cloud CDN: Built on the cloud computing infrastructure, content providers can directly rent CDN services from cloud service providers or lease resources from several SPs to customize their own CDN system [8] [9]. The salient feature of cloud CDN is cost efficiency inherited from the on-demand service paradigm. As such, there are great opportunities to utilize the cloud pricing model to reduce the operational cost, while guaranteeing the service quality: 1) In general, cloud service providers offer three major types of virtual machines (or instances) with different pricing mechanisms, including reserved, on-demand, and spot. For reserved instances, different reservation terms (1-year or 3-year) many incur different levels of pricing. The on-demand instances allow users to hourly pay. In general, the average price of on-demand instances is higher that of reserved instances with same capacity. We can choose different leasing combinations to lower operational cost; 2) Different cloud service providers have distinctive geographical coverage and average prices. In some cases, we need to rent resources from several cloud service providers. We should jointly consider the price of virtual machines, as well as the traffic among different cloud service providers. For example, a cloud provider charges much more for the outgoing traffic to a different cloud provider than inside the same cloud.

SDN or NFV enabled CDN: Recently, software defined networking (SDN) and network functions virtualization (NFV) are proposed to enhance the agility and programmability of networks. SDN separates the control plane and data plane from the traditional network design philosophy. The control plane decides how to handle the traffic, and the data plane forwards traffic according to decisions made by the control plane. SDN adopts a centralized control mechanism to collect network states and make a better decision in comparison with the traditional distributed mechanisms. Furthermore, NFV introduces the virtualization idea to the entire classes of network elements. A straightforward idea is to integrate the conventional CDN with SDN or NFV to increase the service quality. For instance, incorporating SDN to the CDN architecture, CDN providers can provide virtual CDN or software defined CDN services to isolate network traffic among tenants, or prioritize the network traffic among different CDN servers located at different regions for real time applications, e.g., live video streaming.

C. Consumers

Typical end devices for users to enjoy multimedia services include TV screen, smart phone, PC, and notepad. Nowadays, people tend to simultaneously use several devices, also known as second screen, for enriched experience, or inter-connect together for a same service. These two service paradigms present new challenges:

- *Second screen*: also referred to multi-screen or companion screen, is any digital device that is used alongside the first screen and enables the user to interact with the

content on the first screen in some manner. Most often, these devices are smartphones, digital tablets or laptops. Second screen functionality is supported by several interconnected devices and applications, which are seamlessly delivered across all screens. Under this paradigm, we can associate media streaming service with online social network service. For example, when a user is watching a TV program, he/she can interact with friends to discuss this program, or acquire the social response from the online social networks [10] [11] [12].
- *Group Watching*: With the help of online social networks, the group watching scenario, i.e., a crowd of people (e.g., friends, family) watching together, is becoming location independent. People can enjoy multimedia services even if they are geographically distributed by utilizing real-time communication channels. However, the synchronization playback among participating users needs to be guaranteed. Asynchronism may lead to an unpleasant viewing experience and may diminish the feeling of togetherness of the users [13].

III. Learning Based System Design and Optimization

Great efforts have been devoted to study the design and optimization of multimedia networks. These approaches typically assume some static or dynamic models (model-based optimization) for the system, and then develop various algorithms to find the optimal solution via simulation. However, we cannot guarantee algorithms that work well on well-characterized test environments will also perform adequately on real systems that are inevitably more complex. This reality calls for an advanced principle for algorithm design. Fortunately, owing to the open data policy and the affordable cost of developing system prototypes, it is much easier for academic researchers to gain large datasets, and exploit the big data analytics [1] methods to build learning-based optimization methods.

Using abundant machine learning algorithms, we can learn the models to capture the system dynamics, which can optimize the system performance. The models can be in the forms of rules to capture the behavior, decision trees to characterize the features, deep networks to represent the functions, and equations to find the parameters. Based on the learning models, we can optimize the system to either minimize the cost or maximize the reward. Usually, we have the resource limitation while satisfying application requirements. Therefore, we can formulate constrained optimization problems and leverage s-tandard optimization techniques to find the optimal/suboptimal solution to tune the system, in response to the dynamics of the system.

Taking the QoS based video distribution over wireless systems [14] as an example, the algorithm design consists of two steps: 1)constructing the QoE model and using this model to formulate the video distribution problem; 2) finding the optimal solution. In the first step, PSNR is widely exploited to measure the QoS of different versions of a video. However, we need to calculate the PSNR offline, incurring heavy computation cost. In contrast, we can exploit the deep learning algorithm to model the relationship between QoS and the size of a sequence of video frames. In this way, we can estimate the QoS online with less computation cost. After this estimation, we can develop various online algorithms to optimize system performance.

IV. Conclusion

This paper aims to discuss the rising and challenging issues in the future multimedia communication networks. Following an end-to-end architecture of multimedia networks, we present emerging applications, technologies, and research problems for each part. After that, we highlight the concept of learning based system design and optimization.

References

[1] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *Access, IEEE*, vol. 2, pp. 652–687, 2014.

[2] Y. Wen, X. Zhu, J. Rodrigues, and C. Chen, "Cloud mobile media: reflections and outlook," 2014.

[3] H. Hu, Y. Wen, T.-S. C. Chua, Z. Wang, J. Huang, W. Zhu, and D. Wu, "Community based effective social video contents placement in cloud centric cdn network," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1747–1752.

[4] Z. Wang, L. Sun, X. Chen, W. Zhu, J. Liu, M. Chen, and S. Yang, "Propagation-based social-aware replication for social video contents," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 29–38.

[5] V. Krishnamoorthi, N. Carlsson, D. Eager, A. Mahanti, and N. Shahmehri, "Quality-adaptive prefetching for interactive branched video using http-based adaptive streaming," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 317–326.

[6] G.-M. Su, Y.-C. Lai, A. Kwasinski, and H. Wang, "3d video communications: Challenges and opportunities," *International Journal of Communication Systems*, vol. 24, no. 10, pp. 1261–1281, 2011.

[7] Z. Wang, W. Zhu, M. Chen, L. Sun, and S. Yang, "Cpcdn: Content delivery powered by context and user intelligence," *Multimedia, IEEE Transactions on*, vol. 17, no. 1, pp. 92–103, Jan 2015.

[8] Y. Jin, Y. Wen, H. Hu, and M.-J. Montpetit, "Reducing operational costs in cloud social tv: An opportunity for cloud cloning," *Multimedia, IEEE Transactions on*, vol. 16, no. 6, pp. 1739–1751, Oct 2014.

[9] Y. Jin, H. Hu, X. Liu, and Y. Wen, "Mutas: Multi-screen tv experience as a service through cloud centric media network," in *Sensing, Communication, and Networking (SECON), 2014 Eleventh Annual IEEE International Conference on*. IEEE, 2014, pp. 146–148.

[10] H. Hu, Y. Jin, Y. Wen, T.-S. Chua, and X. Li, "Toward a biometric-aware cloud service engine for multi-screen video applications," in *Proceedings of the 2014 ACM conference on SIGCOMM*. ACM, 2014, pp. 581–582.

[11] H. Hu, J. Huang, H. Zhao, Y. Wen, C. W. Chen, and T.-S. Chua, "Social tv analytics: a novel paradigm to transform tv watching experience," in *Proceedings of the 5th ACM Multimedia Systems Conference*. ACM, 2014, pp. 172–175.

[12] H. Hu, Y. Wen, H. Luan, T.-S. Chua, and X. Li, "Toward multiscreen social tv with geolocation-aware social sense," *MultiMedia, IEEE*, vol. 21, no. 3, pp. 10–19, July 2014.

[13] B. Rainer and C. Timmerer, "Self-organized inter-destination multimedia synchronization for adaptive media streaming," in *Proceedings of the ACM International Conference on Multimedia*. ACM, 2014, pp. 327–336.

[14] A. Testolin, M. Zanforlin, M. De Filippo De Grazia, D. Munaretto, A. Zanella, M. Zorzi, and M. Zorzi, "A machine learning approach to qoe-based video admission control and resource allocation in wireless systems," in *Ad Hoc Networking Workshop (MED-HOC-NET), 2014 13th Annual Mediterranean*. IEEE, 2014, pp. 31–38.