Efficient Cloud Provisioning for Video Transcoding: Review, Open Challenges and Future Opportunities

Maria G. Koziri[1], **Panos K. Papadopoulos**[2], **Nikos Tziritas**[3], **Thanasis Loukopoulos**[2], **Samee U. Khan**[4], **Albert Y. Zomaya**[5]

[1]Department of Computer Science, University of Thessaly,
2-4 Papasiopoulou str., P.O. 35100 Lamia, Greece, mkoziri@uth.gr

[2]Department of Computer Science and Biomedical Informatics, University of
Thessaly, 2-4 Papasiopoulou str., P.O. 35100 Lamia, Greece,
{ppapadopoulos,luke}@dib.uth.gr

[3]Research Center for Cloud Computing, Chinese Academy of Sciences, Shenzhen
Institutes of Advanced Technology, Shenzhen, China, nikolaos@siat.ac.cn

[4]Electrical and Computer Engineering, North Dakota State University,
Fargo, USA, samee.khan@ndsu.edu

[5]School of Information Technologies, University of Sydney, Building J12, Sydney,
NSW 2006, Australia, albert.zomaya@sydney.edu.au

Abstract

Video transcoding is the process of encoding an initial video sequence into multiple sequences of different bitrates, resolutions and video standards, so that it can be viewed from devices of various capabilities and network access characteristics. Since video coding is a computationally expensive process and the amount of videos in social media networks drastically increase every year, large media providers' demand for transcoding Cloud services will continue rising. In this article we survey the current state of the art on related Cloud services. We also summarize research on video transcoding and provide indicative results for a transcoding scenario of interest related to facebook. Finally, we illustrate open challenges in the field and outline paths for future research.

Keywords: video compression, video coding, transcoding, Cloud, HEVC, facebook.

Overview

As reported by Cisco in its mobile data traffic forecast for the years 2015-2020, mobile Internet traffic increased by 74% during 2015, with 55% of it due to video. Given the popularity of social media, the affordable prices of smart devices with high resolution cameras and the everlasting user need for social interaction and entertainment, video traffic on the Internet, particularly the mobile one, will continue rising. To further complicate things, large media providers have to satisfy clients viewing video from a plethora of different devices with various players and capabilities that reside behind network connections of various speeds. This necessitates video transcoding which is the process of encoding an initial video sequence into one or more sequences of different resolution level, bitrate, quality and perhaps coding standard, e.g., from H.264/AVC to HEVC.

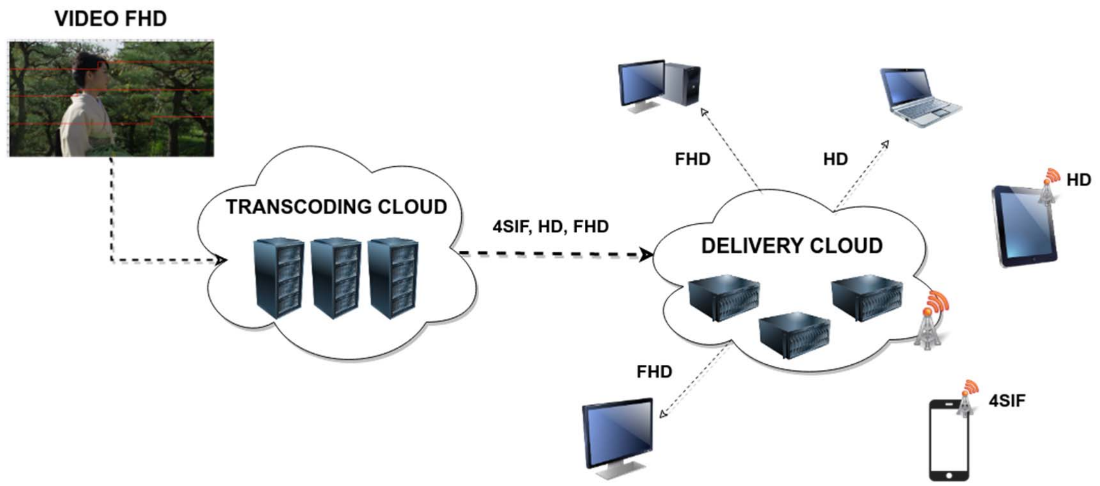

Figure 1. A generic Cloud transcoding architecture. The initial FHD video sequence is sent to the transcoding Cloud that produces output sequences in 3 different resolutions. These sequences are sent for delivery to various devices.

In its simplest form, transcoding can be done by decoding the original video and re-encoding it using the desired parameters. Since video encoding is a computationally intensive task and given that the amount of videos that must be transcoded before delivery can exceed the infrastructure capacity of most media providers, using Cloud resources seems to be the only scalable solution. Fig. 1 shows schematically a generic Cloud transcoding scheme.

Transcoding as a Service

A straightforward approach to use Cloud resources for transcoding, is the IaaS model whereby the user should estimate the computational resources required for the tasks, select a codec to use, create, schedule and monitor job batches as well as manage the output streams. As it is apparent, this solution entails a non-negligible overhead in specialized personnel costs.

This is the reason why recent years have witnessed a growth to the number of companies offering PaaS or SaaS solutions, effectively implementing Transcoding as a Service (TaaS). TaaS targets at reducing the complexity from a client's point of view in: (i) defining transcoding tasks, (ii) executing them, and (iii) managing their output. Concerning (i) the client is offered with predefined encoding modes either optimally tailored for specific devices or tailored for particular resolutions. Such predefined modes can be of huge help to small scale clients who can't afford the complexity of defining optimal encoding parameters, e.g., bitrate, QP (Quantization Parameter) for their targeted resolutions, while large media providers can still use their home settings ignoring the predefined modes. As far as (ii) is concerned, the client neither is required to know the details of the codec used, nor is he required to schedule its tasks. Finally, concerning (iii) TaaS providers can forward videos to a Content Delivery Network after packaging them so that they are ready for streaming, e.g., in MPEG-DASH or HLS. Most of the providers also offer watermarking and DRM (Digital Rights Management) capabilities as well as options for adding advertisements.

TABLE I. CLOUD TRANSCODING PROVIDERS

[1]Company	codecs	Live transcoding	SLAs	Pricing	[2]File encoding scenario	[3]Live stream encoding scenario
brightcove ZENCODER	H.264/AVC, VP8, VP9, Theora, HEVC	YES	YES	Output duration, Live channel duration	360$	4,500$
encoding.com	H.264/AVC, VP8, Theora, HEVC	YES	YES	Output size, Per channel flat price	~1,250$	899$
Amazon Elastic Transcoder	H.264/AVC, VP8, VP9	NO	NO	Output duration	180$	N/A
telestream | cloud	H.264/AVC, VP8, VP9, Theora, HEVC	NO	YES	Output duration	239$	N/A
Wowza Streaming Cloud	H.264/AVC, HEVC	YES	NO	Processing time, Output size, Per channel rate, Number of streams, Stream duration	N/A	999$
Microsoft Azure Media Services	H.264/AVC, Theora	YES	YES	Output size, Live channel duration	180$	1,311$
BITMOVIN	H.264/AVC, VP9, HEVC	YES	NO	Output size	~3,305$	~10,555$

1. The information is based on companies' Web sites as of April 2017. In the price charging scenarios the most economic package was selected. Prices are for coding; other charges such as storage, transmission etc. may apply.
2. The scenario consists of transcoding 100 hourly length videos, into 100 FHD output sequences. Each output sequence was estimated to be 25GB.
3. The scenario consists of setting a live stream. The input stream is assumed to be transcoded into a single output stream of FHD quality. The total channel duration considered was 300 hours monthly.

A distinguishing factor among the available solutions is whether they offer live video transcoding or not. Elastic Transcoder by Amazon and telestream|cloud offer file transcoding without supporting live streaming. On the other hand Wowza streaming Cloud targets live transcoding and streaming. Another distinguishing factor among TaaS providers are the pricing schemes they use. In case of file transcoding solutions, providers usually charge either the output video duration or the output video size, whereas if live video transcoding is included in the package, the consumed bandwidth or fixed per channel charge is also added to the bill. Table I summarizes some basic characteristics of TaaS providers and gives estimated charging costs for two scenarios, one involving live transcoding and another stemming from VoD. We would like to mention that the table is not intended to provide a selection list since it neither includes all possible providers, nor pricing is the single decision criterion. Furthermore, prices usually don't scale linearly to workload size, while they might also be negotiable (large customers).

The case of SVC

A problem with transcoding is that a single video sequence leads to multiple output files. SVC (Scalable Video Coding) aims at solving it, by coding in one bitstream multiple fidelity points of the original video signal. The fidelity points can be temporal resolution, spatial resolution, or quality (Signal to Noise Ratio-SNR). Thus, the bitstream is comprised of layers, i.e., subsets of the bitstream that represent the additional information needed to be able to decode the signal at an increased quality. Layers form a hierarchy whereby a specific layer needs all lower ones it refers to, in order to be decoded. Therefore, each layer, in combination with the layers it depends on, forms a representation of the video signal in a specific spatial-time resolution and quality. The interested reader is referred to [1] for a detailed overview of SVC extensions in H.264/AVC. SVC has seen some successful deployment in the industry, e.g., Vidyo's conferencing service. Nevertheless, it hasn't replaced transcoding due to lack of existing popularity and interoperability difficulties. Additionally, transcoding is necessary when changing video standard.

Research on Cloud Transcoding

Most TaaS providers don't give details concerning algorithms and system architecture design. For this reason in the article we focus on surveying research from academia.

Efficient transcoders

A significant amount of work exists on how to efficiently implement transcoders. When the requirement is to produce a video sequence of the same format but presumably of a lower bitrate (also called transrating) a straightforward yet inefficient approach is to use an open loop method whereby the decoded sequence is re-encoded using larger QP values. Closed loop methods involving error compensation were also proposed. A survey on transrating can be found in [2]. Of particular interest is the case where a sequence must be transcoded on different video standards.

In [3] the authors propose a scheme for transcoding HEVC to VP9. Through exploiting information from HEVC decoding process (mainly the prediction modes Intra/Inter and the reference frames) they prune decisions of VP9 encoder, improving time up to 60%.

In [4] an H.264/AVC to HEVC transcoder is presented. The proposed scheme achieves a speedup of 7.89x compared to full re-encoding, while the coding efficiency loss is 3.28% BD-Rate. This speedup is achieved by two modifications in the HEVC encoder. First, a fast mode decision framework based on a post-order traversal of the CTU (Coding Tree Unit) quadtree is proposed. This framework in combination with information from the H.264/AVC sequence and a modified RD (Rate Distortion) cost prediction model achieves early termination of the mode decision process. Secondly, a new fast motion estimation algorithm is implemented that selects the best candidate from a list of previously encoded H.264/AVC and HEVC motion vectors.

Another H.264/AVC to HEVC transcoder is presented in [5]. Once again the target is to accelerate the mode decision process in HEVC by using information available in

H.264/AVC sequence. This is done by a Fast Quadtree Level Decision (FQLD) algorithm. FQLD exploits the information gathered at the H.264/AVC decoder to decide on CU (Coding Units) splitting in HEVC using Naïve-Bayes probabilistic classifier. Results show that a speedup of up to 3.98x is achievable without significant RD loss.

System design and scheduling of transcoding jobs in the Cloud

In [6] a Map-Reduce approach for video encoding is described. The method is based on splitting an initial sequence into chunks, whereby chunk size is a multiple of GOP (Group Of Pictures) size, process each chunk separately, and merge them in a final step to produce the final compressed video sequence. Instead of having fixed number of GOPs per chunk, the authors of [7] proposed to adapt chunk size. The premise is to avoid breaking dependencies (at chunk boundaries) between GOPs of the same scenery.

In [8] a Cloud based transcoding system is presented. The system architecture consists of: (i) task manager that accepts user requests and checks whether they can be satisfied by the cached files, (ii) transcoding servers, (iii) downloaders which are responsible for locating and fetching a video that isn't already cached and (iv) a task dispatcher controlling the rate of downloading and transcoding activities. A significant part of transcoding tasks is performed during light load periods (e.g., night time) as a prefetching strategy.

In [9] the authors propose a system architecture that enables video encoding at the edges of a mobile network. They showed that energy savings in mobile devices can be achieved by shifting some of the encodings towards edge servers. The core of their proposal is to tune the sending devices to encode videos at a high bitrate, fast, thus, saving energy (but consuming larger bandwidth). The videos will then be transcoded at the edges for final delivery.

The focus of [10] is to reduce the computational load of transcoding. For each video (or the popular ones) the first chunk of it is transcoded in an offline manner to reduce buffering. The remaining chunks are transcoded in an online manner using a Markov chain estimator that decides which video parts will be needed once downloading starts. Partial transcoding is also the focus of [11] whereby the aim is to optimize the long term operational cost of a video delivery service. Such cost comes in terms of storage by caching transcoded video segments and in terms of computational effort due to transcoding uncached segments. An online algorithm is proposed that decides which segments should be cached based on Lyapunov optimization and queuing theory. It is shown that operational cost can be reduced by 30%. An admission control algorithm for transcoding requests arriving at a cluster is proposed in [12]. The algorithm might choose to accept a transcoding job if servers' working queues are lightly loaded, or defer on the request, redirecting it to an entertainment server. Deferred requests are rejected when the entertainment server becomes overloaded or accepted in the next admission slot if server queue length allows it.

Live video transcoding was considered in [13], focusing on the case where sources don't have strict high QoS demands. The authors analyzed datasets from Twitch concerning live video casting and tackled the problem of optimizing user experience.

Perfect user experience is achieved when transcoding to the maximum user supported resolution and bitrate. An ILP formulation was developed to decide for every stream the formats to be transcoded, given existing hardware.

Research summary and discussion

The aforementioned research, targets at optimizing the operation of transcoding Clouds by:

- Speeding up the transcoding time of a single task.
- Achieving predefined quality levels.
- Achieving real-time or SLA based performance.
- Scheduling multiple transcoding tasks optimally.
- Reducing resource and energy consumption.

These goals are often conflicting with each other hence, there is a need for further research to develop methods that achieve better trade-offs.

Future challenges

A number of future challenges exist, especially as more and more of video traffic will concern 4K sequences. Following, we present some of them.

Efficient Cloud resource management

Research on Cloud transcoding resource management is far from being a closed topic, especially as energy consumption and network overhead become increasingly important. The need for holistic approaches both at system level design and algorithmic concepts that tackle in unified way different transcoding scenarios, e.g., standalone files, batch of VoD transcoding tasks, live casting etc., is urgent as the relevant market can’t be overlooked by TaaS providers. Such unified approaches should also include SVC (wherever applicable). It is also apparent that research exploiting the potential of edge computing in minimizing network and computational overhead will be of paramount importance in an era of 4K and 8K resolutions.

Per sequence transcoding ladder

In most cases media providers fix the bitrates and resolutions for transcoding (also termed encoding/transcoding ladder). As advocated in [14] by Netflix, the one size fits all is not necessarily the best approach. In fact it was found that movies have different bitrate demands depending on category. For instance, cartoons could be compressed with reasonable quality using less than half the bitrate of other movies. Research on the area especially when combined with user perceived quality metrics promises to reduce the bitrates of encoded videos without affecting quality.

Adoption of newer video standards

Increased compression efficiency at the same quality is the premise behind any newly developed video standard. HEVC together with the anticipated AV1 and the planned

JVET will form the state of the art on video coding standards in the coming years. Assuming that the royalties problem that stalled HEVC adoption is resolved, or AV1 arrives delivering its performance claims, there is going to be a surge in transcoding demand in the foreseeable future.

The scale as well as the costs involved in the above endeavour could be massive. YouTube experiences a workload of >300 hours of uploaded videos per minute. Assume that all the videos uploaded within a year must be transcoded in a new standard, with one output sequence per input file. By using the pricing scheme of Amazon's Elastic Transcoder the total financial cost will be between 142M and 284M. In case more output sequences are needed, the budget can scale in the order of billions. Computational demands will also be huge. Assuming the assigned VMs achieve real time performance (24 fps), 18,000 VMs working 24/7 for a whole year must be allocated.

From a media provider's point of view, research on efficiently using the available Cloud resources to maintain QoS in a cost effective manner, will be of primary importance. This will likely entail careful selection of the videos to be transcoded. As it is illustrated in the sequel, for a popular social media network, this can be achieved at relatively small cost.

Facebook experiment

Here we characterize the impact on facebook traffic by moving from H.264/AVC to HEVC. We used the x264 and x265 codecs respectively for H.264/AVC and HEVC with PSNR-tuned coding settings suggested by [15]. Experiments were run over 4 identical servers each with two 6-core Intel Xeon E5-2630 CPUs at 2.3GHz and 256GB memory.

Dataset

To the best of our knowledge no suitable dataset existed for the experiment we wished to conduct, therefore we resorted to create our own. We collected 200 facebook videos by selecting the top-20 videos from each of the top-10 video publishers as they appear in the rankings of May 2016 in `tubularlabs.com`. We downloaded each video in the highest available resolution which for 168/200 videos was 720p or more. All videos were in `mp4` format. For the purpose of the experiment transcoding was performed by decoding each video using x264 codec and re-encoding it in HEVC using x265. Table II details video publishers, dataset characteristics and the results of the transcoding experiment.

Before proceeding to the results we discuss how well the dataset characterizes facebook traffic using the following observations: (i) The total number of views for the top-10 publishers on May 2016 was 9,304M. By using the average video size calculated in the dataset (59.92 secs), these views translate into ~5M hours of video views per day; (ii) Our dataset accounts for 36.6% of this volume; (iii) As of January 2016 it was reported (`techcrunch.com`) that an estimated 100M hours of video was viewed daily on facebook. Even if the explosive growth of facebook means this value probably became much larger on May 2016, the top-10 publishers still account for a considerable ratio (5% with Jan. 2016 statistics), of which the dataset accounts for roughly one third. Therefore we can state that the created dataset captures a sufficient portion of facebook video traffic.

TABLE II. TRANSCODING A FACEBOOK DATASET INTO HEVC

	Publisher Statistics			
Names	Tasty, UNILAD, The LAD Bible, Viechten met Daan, NowThis, Tastemade, FailArmy, CH51, PlayGround Video, Nifty			
Categories	Food (2), Entertainment (4), News(2), Style & Beauty (1), Home & DIY (1)			
Total views in May 2016 (e^6)	9,304			
	Dataset Statistics			
Number of videos	200 (top 20 per publisher)			
Resolutions	23 different resolutions, 720p (82 videos), 1080p (62 videos)			
Total views (e^6)	3,410			
Average views (e^6) per video	17.04			
Total video size (MB)	1,352.66			
Average video size (MB)	6.76			
Total video duration (secs)	11,984			
Average video duration (secs)	59.92			
Total number of frames	325,962			
Total network overhead (TB)	24,509.52			
	Transcoding Results			
	Total Video Size (MB)	**Total Network Overhead (TB)**	**Average PSNR**	**Aggregate Core Utilization Time (secs)**
H264/AVC to HEVC	708.21 (47.64% reduction)	11,435.19 (53.34% reduction)	44.35	14,272,397

Results

The experiment attempts to answer the following two questions:
- *What would the gains be if instead of H.264/AVC, HEVC was used?*
- *What is the related computational and financial overhead?*

To answer the above, dataset videos were transcoded into HEVC using their original resolutions. Results indicate that compared to H.264/AVC the aggregate video size is reduced by 47.64%. Network overhead is calculated assuming that all views were for the maximum resolution and by counting only video file sizes (the extra cost of network packaging is rather negligible). Table II records a reduction of 53.34% compared to the original sequences, while PSNR remained high.

Runtime results depict the aggregated execution times of the transcoding tasks when using one thread each. In order to finish the 200 transcodings, the 4 available 12-core servers should work continuously at maximum core utilization (12 tasks per server) for time duration of 3.44 days. The C3 EC2 instance is suggested by Amazon for video encoding. Since the PassMark ratio between E5-2680 CPU (c3.8xlarge) and E5-2630 (our servers) is 2.2, it is expected that (all other things being equal) running the tasks on 4 equivalent C3 instances would require ~1.56 days, for a spot starting cost of 59.9$.

Bearing in mind that our dataset accounted for roughly 1.8% of the monthly total view load, results indicate that social media networks tailored for exchanging small duration videos will benefit dramatically by moving to a newer standard. These benefits can materialize by transcoding only a small portion of the most popular videos at low prices using TaaS providers.

Acknowledgments

Samee U. Khan's work was supported by (while serving at) the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Nikos Tziritas' work was supported by NSFC and PIFI International Scholarship under the grants 61550110250 and 2017VCT0001, respectively.

References

[1] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of Scalable Video Coding Extension of H.264/AVC Standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17(9), pp. 1103–1120, Sep. 2007.

[2] I. Ahmad, X. Wei, Y. Sun, and Y.-Q. Zhang, “Video Transcoding: An Overview of Various Techniques and Research Issues,” in IEEE Trans. Multimedia, vol. 7(5), pp. 793-804, Oct. 2005.

[3] E. de la Torre, R. Rodriguez-Sanchez, and J. L. Martínez, “Fast Video Transcoding from HEVC to VP9,” in IEEE Transactions on Consumer Electronics, vol. 61, no. 3, pp. 336-343, Aug. 2015.

[4] J. F. Franche and S. Coulombe, “Fast H.264 to HEVC Transcoder based on Post-Order Traversal of Quadtree Structure,” in Proc. 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 477-481.

[5] A. J. Díaz-Honrubia, J. L. Martínez, J. M. Puerta, J. A. Gámez, J. De Cock, and P. Cuenca, “Fast Quadtree Level Decision Algorithm for H.264/HEVC Transcoder,” in Proc. 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 2497-2501.

[6] R.S. Pereira and K. Breitman. Video Processing in the Cloud. SpringerBriefs in Computer Science, ISBN 978-1-4471-2137-4, 2011.

[7] M.R. Zakerinasab and M. Wang, “Dependency-Aware Distributed Video Transcoding in the Cloud,” LCN 2015, pp. 245-252.

[8] Z. Li, Y. Huang, G. Liu, F. Wang, Z.-L. Zhang, and Y. Dai, “Cloud Transcoder: Bridging the Format and Resolution Gap between Internet Videos and Mobile Devices,” in Proc. NOSSDAV 2012, pp. 33-38.

[9] M.T. 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 Proc. ICIN 2015, pp. 38-44.

[10] D.K. Krishnappa, M. Zink, and R.K. Sitaraman, “Optimizing the Video Transcoding Workflow in Content Delivery Networks,” in Proc. MMSys 2015, pp. 37-48.

[11] G. Gao, W. Zhang, Y. Wen, Z. Wang, and W. Zhu, “Towards Cost-Efficient Video Transcoding in Media Cloud: Insights Learned From User Viewing Patterns,” IEEE Trans. Multimedia, vol. 17(8), pp. 1286-1296, Aug. 2015.

[12] A. Ashraf, F. Jokhio, T. Deneke, S. Lafond, I. Porres, and J. Lilius, “Stream-Based Admission Control and Scheduling for Video Transcoding in Cloud Computing,” in Proc. CCGRID 2013, pp. 482-489.

[13] R.A. Pardo, K. Pires, A. Blanc, and G. Simon, “Transcoding Live Adaptive Video Streams at a Massive Scale in the Cloud,” in Proc. MMSys 2015, pp. 49-60.

[14] A. Aaron, Z. Li, M. Manohara, J. De Cock, and D. Ronca, “Per Title Encode-Optimization”, Netflix Tech blog, 14 Dec. 2015. Available at: http://techblog.netflix.com/2015/12/per-title-encode-optimization.html

[15] J. De Cock, A. Mavlankar, A. Moorthy, and A. Aaron, “A Large-Scale Video Codec Comparison of x264, x265 and libvpx for Practical VOD Applications,” Proc. of SPIE, Applications of Digital Image Processing XXXIX, vol. 9971, 997116, Sept. 2016.

Maria G. Koziri received a PhD degree in Computer Science from the University of Thessaly (UTH), Greece, in 2007. She is currently a Visiting Lecturer in the Computer Science Dept., UTH. Her research interests are in the field of video compression: codec optimization, hardware design, scalable video coding.

Panos K. Papadopoulos received his BS degree by the Computer Science and Biomedical Informatics Dept. of the Univ. of Thessaly, Greece in 2015. He is currently a PhD student in the same department. His research interests include: video coding and cloud computing.

Nikos Tziritas received his PhD degree from the University of Thessaly, Greece, in 2011. He is currently a researcher at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include scheduling, load balancing and replication in CDNs as well as energy optimization and resource management in WSNs and cloud computing systems.

Thanasis Loukopoulos is Assistant Professor at the Department of Computer Science and Biomedical Informatics, University of Thessaly, Greece. His research interests include: green, cluster, grid and cloud computing. His work appeared in over 50 publications.

Samee U. Khan is Associate Professor of Electrical and Computer Engineering at the North Dakota State University, Fargo, ND, USA. His research interests include optimization, robustness, and security of: cloud, cluster and big data computing, social networks, wired and wireless networks, power systems, and optical networks. His work has appeared in over 300 publications. He is a Fellow of IET, BCS and a Senior IEEE member.

Albert Y. Zomaya is the Chair Professor of High Performance Computing & Networking in School of Information Technologies, University of Sydney. Prof. Zomaya is the author/co-author of seven books, more than 400 papers, and the editor of 12 books and 15 conference proceedings. He was the Editor in Chief of the IEEE Transactions on Computers and serves as an associate editor for 19 leading journals. He is a Fellow of IEEE, IET and AAAS.