Epidemic Models using Resource Prediction Mechanism for Optimal Provision of Multimedia Services

Yiannos Kryftis, Constandinos X. Mavromoustakis
Department of Computer Science
University of Nicosia
Nicosia, Cyprus
kryftis.y@unic.ac.cy, mavromoustakis.c@unic.ac.cy

Jordi Mongay Batalla
National Institute of Telecommunications
Szachowa Str. 1,
04-894 Warsaw, Poland
jordim@interfree.it

George Mastorakis
Department of Informatics Engineering
Technological Educational Institute of Crete
Heraklion, Crete, Greece
gmastorakis@staff.teicrete.gr

Periklis Chatzimisios
CSSN Research Lab,
Alexander TEI of Thessaloniki,
Thessaloniki, Greece
peris@it.teithe.gr

***Abstract*—This paper proposes a network architecture that goes a step beyond the current state-of-the-art, by elaborating on a novel resource reservation and provision scheme, through a Media Distribution Middleware (MDM), that is based on a Resource Prediction Engine (RPE) leading to both high resource utilization and quality guarantees. The architecture defines management planes and software components that provide the mechanisms for collecting monitoring data, predicting possible future values of the network metrics and resources usage, and applying management decisions to keep the provision optimal. The proposed research approach is based on novel time series and epidemic spread models, and the outcome is used for the optimal distribution of streaming data, among Content Delivery Networks, cloud-based providers and Home Media Gateways. The proposed epidemic diseases model adopts the characteristics of the multimedia content delivery over the network architecture. In this context, the paper aims to present the advantages of using such models, by presenting and analyzing an epidemic spread scheme for Video on Demand (VoD) delivery, to predict future epidemic spread behavior. The validity of the proposed system is verified through several sets of extended experimental simulation tests, carried out under controlled simulation conditions.**

Keywords: Resource Prediction Engine, Content Delivery Networks, Multimedia Services Systems, Quality of Experience, Epidemic Models

I. INTRODUCTION

The research and innovation in the field of multimedia content distribution is significant during the few years to support the increasing demand for efficient media distribution over the Internet. The multimedia services distribution generates today a significant percentage of the whole IP traffic and based on [1] this number is expected to be further increased by 2018 leading to the Future Media Internet. The advances in social networks contribute to this direction by providing to the users an environment that rapidly evolves giving them easier access to more multimedia content. The existing cloud infrastructures that utilize the virtualization of resources, and the Next Generation Networks that provide ways of clever management for the distribution of resources are the keystones of the upcoming Future Media Internet. An important challenge in providing a high Quality of Experience (QoE) to the user is the ability to exploit optimal provision of multimedia services. Tackling such challenges, this paper goes beyond the state of the art, elaborating on novel multimedia services delivery solution that utilizes a resource prediction engine and innovative models to achieve the optimum allocation of the resources used for content transmission. This is done through the exploitation of existing servers infrastructures capabilities available in conventional clouds (i.e. public or private computing infrastructure configurations, usually offered by over-the-Top providers), in Content Delivery Networks (CDNs) and in Home Media Gateway Clouds (i.e. Home Gateways/Community Gateways configurations, exploited in peer-to-peer mode). The resource usage prediction is based on statistical models able to accurately predict the possible future values given some historical data, and on a novel epidemic model able to forecast a possible upcoming huge increase in the demand of a specific resource. This is the case where a resource like a Video on Demand (VoD) becomes viral through social media and the interaction among users and there is an enormous computational and network need to serve the VoD. Following this introductory section, section II presents the related work and the research motivation of this paper. Section III elaborates on the proposed research approach based on a novel network architecture and a resource prediction engine for optimal multimedia services provision. Finally, section IV provides the performance evaluation results and section V concludes the paper, by indicating future research.

II. RELATED WORK AND RESEARCH MOTIVATION

There is a plethora of research works on the combination of different delivery methods, in order to achieve better QoE for the users [2]. In [3] Xu et al. propose a streaming media distribution architecture that combines the advantages of using CDN for providing high QoE with the low cost of using P2P-based stream. Yin et al. in [4] present the design and deployment of a Hybrid CDN-P2P System for Live Video Streaming, showing that the combination improves the startup delay time and stability. Current research approaches consider mainly on how to overall benefit from the combination of the different delivery methods but they don't handle each resource separately. This paper goes beyond the current state-of-the-art, by handling each resource (i.e. streaming channel) separately based on current and forecast demand for each resource, as well as the predicted network metrics.

The Resource Prediction Engine (RPE) is essential for the proposed multimedia content delivery system, in order to provide the desired QoE to the end users. It is responsible for the prediction of the needed bandwidth capacity and other network metrics based on previously collected data. The prediction engine has to be based on novel methods and models that can accurately forecast the future demands, in order to trigger through a management plane the proper actions for keeping the desired quality for the streaming sessions. In [5] Niu et al. present some time-series analysis techniques to predict the server bandwidth demand and the peer upload for content delivery in peer-assisted Video-on-Demand (VoD) services. The expectations for future demands are calculated based on the history of demands as monitored by cloud monitoring services. For the prediction of future population of each video channel, they utilize the seasonal ARIMA (auto-regressive integrated moving average) model [6], for avoiding the periodicity. They infer the initial population of a new released video based on machine learning techniques and past data from newly released channels as training data. For the prediction of the server bandwidth demands by a video channel at future time, the ARMA (auto-regressive moving-average) model was used [6].

Social networks through Internet play a significant role in content delivery, by providing ways of interactions among users that can lead to a lightning spread of content [7]. It is essential to take into account the recent advances in connected media technologies. In [8] Gonccalves et al. suggested a probabilistic resource provisioning approach that utilizes epidemic models to represent sudden and intense workload overflow in VoD delivery process. There is a plethora of bioinspired models used to improve aspects such as handling large-scale networks [9]. Meloni et al in [10] present the epidemic spread in scale free networks with emphasis on the spread of computer viruses that presents similar characteristics to the epidemic spread of human diseases. In this context, this paper extends the basic disease model, and presents a novel model that can be used to accurately describe the multimedia content delivery (i.e. VoD delivery).

The vast majority of disease models are based on a splitting in compartments of the individuals in a population based on their disease status [11]. A basic epidemic model is the Susceptible-Infectious-Recovered (SIR) model used as the basis for most others. The differential equations that describe the model are the following:

$$\frac{dS}{dt} = \frac{\beta.S.I}{N} \tag{1.1}$$

$$\frac{dI}{dt} = \frac{\beta.S.I}{N} - \gamma.I \tag{1.2}$$

$$\frac{dR}{dt} = \gamma.I \tag{1.3}$$

$$S + I + R = N \tag{1.4}$$

In these equations, S, I and R refer to the number of susceptible, infectious and recovered individuals, respectively, in a population of size N. The other parameters are the effective number of contacts per unit time (β), and the mean infective period ($\frac{1}{\gamma}$). The nonlinear term $\frac{\beta.S.I}{N}$ represents the transmission of infection that generates a variety of rich dynamical behaviours. Theoretical modelling of how diseases spread in complex networks is based on the assumption that the propagation is driven by reaction processes and that the transmission occurs from every infected neighboring entity at each time step, producing a diffusion of the epidemics on the network. This paper presents an epidemic model suitable for the description of content delivery and elaborates on a network architecture that predicts the future content delivery demands and the future network usage, by utilizing novel models and algorithms, performing all the necessary adaptations to deliver the content in an optimal and balanced way.

III. UTILIZATION OF A RESOURCE PREDICTION ENGINE AND EPIDEMIC MODEL FOR OPTIMAL MULTIMEDIA SERVICES PROVISION

Towards introducing a Resource Prediction Engine that utilizes Epidemic Models, for the provision of multimedia services, there is a need of a novel architecture to provide a collaboration environment with management components in separate layers that cooperate during the multimedia delivery process [12]. The proposed network architecture consists of the following components distributed in the Management Planes as shown in Fig1. The Media Distribution Middleware (MDM) and the Media QoE Meter (MQM) are placed in Delta M&C plane located in the premises of the content provider (possibly an ISP provider), the Media Services Manager (MSM) is placed at the CDN and Cloud M&C plane, the Media Advanced Streamer (MAS) is positioned next to the actual streaming CDNs and Cloud Servers. Finally the Enhanced Home Gateway (EHG) are part of the home equipment of the users.

The control modules of EHGs constitute the Media Home Gateway Cloud (MHGC) M&C plane and they are responsible for creating the MHGC ad-hoc system from a set of peer-to-peer connected EHGs. Each EHG receives content requests from the users, requesting data from the MDM, about which MHGC peers should get involved to efficiently deliver requested content. EHG collaborates with MSM entities that reside in CDN/Cloud M&C planes and manage all Service Provider's resources, to obtain media content requested by the user, if the content is not stored on any of EHGs belonging to given MHGC. The MSM, according to the recommendations received from the MDM, takes a decision, on which server should stream the requested media and with which bitrate. In this way, the MSM, contrary to the existing solutions, performs adaptation decision, taking into account not only the available bandwidth, but also considering other important information addressed by the MDM, such as the estimated QoE value and the prediction of the potential upcoming streaming sessions.

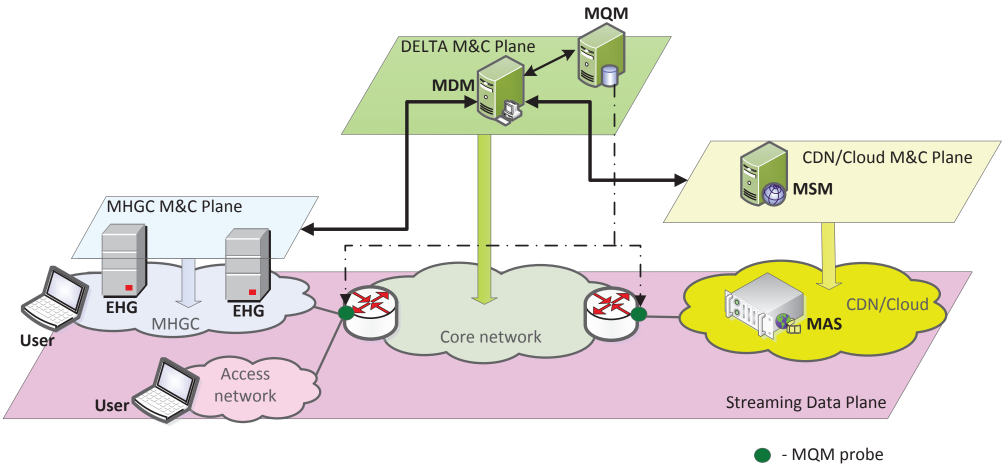

Fig. 1. Proposed Network Architecture

The MAS entity resides in the CDN/Cloud domain as a standalone component. Its role is to perform the streaming process, according to the instructions received from the MSM/EHG entity. MQM component is responsible for continuous monitoring of network metrics at the users and the Service Provider's domain access points, as well as the users' context and preferences. Based on the data gathered by the set of the MQM probes, distributed all over the domain, this entity provides to the MDM the related data about the current network conditions and the estimated value of QoE available for a user. Moreover, the MQM sends alerts to the MDM, only if any of the monitored QoS/QoE parameters declines below the allowed level. MDM is the main component of the M&C plane. It executes all necessary operations and determines all data required for optimal allocation of the available resources at each Resource Provider's domain. As a result, the MDM returns guidelines, which resources should be used for handling given user's request, to achieve the best (in terms of efficiency) resource exploitation. The MDM adopts a resource prediction engine, in order to be able to predict future demands for resources. The prediction takes as input the demand for each resource in the past, using a combination of statistical methods and algorithms for the adaptivity to description models, in order to predict future demands. This provides the opportunity to the system to make the optimal distribution of data in Clouds, CDNs and EHGs based on the prediction before the actual need. Fig. 2 presents the internal architecture of the MDM component. The QoS/QoE Politics Traffic Data History component gathers the monitoring data that comes from the MQM, using it as input to the Media Traffic Forecast and generating the prediction for the traffic in the network. The Media Traffic Forecast utilizes the prediction epidemic model for the prediction of the upcoming epidemic or not spread of the content and the time series models for the prediction of future values of specific metrics for the resources. The time series models utilized for the prediction are based on ARIMA models, exponential smoothing methods, Theta method [13], cubic splines [14] and many others. More details about the models and algorithms are presented on our previous publications [15], [16]. The methodology for the prediction based on time series models can be summarized as the utilization of monitoring metrics (i.e bandwidth usage) from the past to match the model that fits and based on that to predict the possible future values. The outcome of the forecasts is used as an input to the Resource allocator/scheduler that uses algorithms to combine the forecast methods, allowing it to decide on the optimal delivery methods. The results are feeding the MSM component with recommendations, on which server should stream the requested media, while at the same time the Bandwidth Allocation Optimizer calculates the optimal bandwidth allocation for the P2P delivery between the MHGC devices. It is an online system, which takes into consideration the network metrics that come from the MQM, delivering that information directly to the MHGC devices. MHGC devices exchange management information between them and together they constitute a M&C plane that manages the P2P network between the EHG devices.

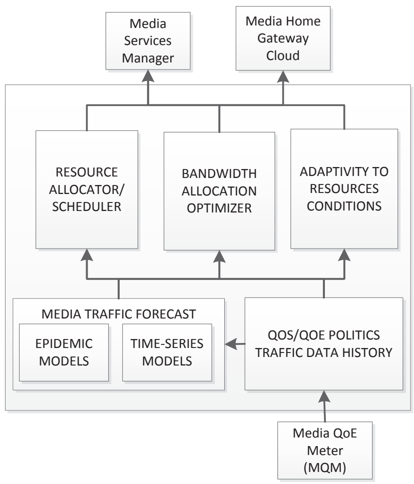

Fig. 2. MDM Internal Architecture

Social networks play a vital role in the generation of traffic and spread of a Video On Demand (VoD) over the Internet. The presented work exams the epidemic models for use for the prediction of VoD usage as part of the general issue of optimal content delivery and it suggests a specific model and methods to predict and proactively handle the cases of enormous spread of content. The proposed model divides the population into several compartments based on the percentage of the population in each state as shown in Fig. 3.

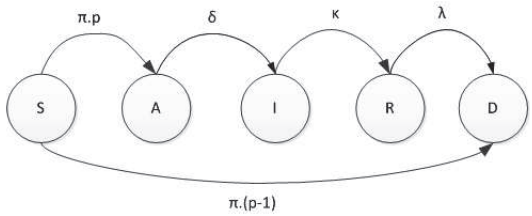

Fig. 3. Proposed epidemic model for resource usage prediction

The Susceptible (S) group includes the subscribers that are eligible to download the Video, the Active (A) includes those that are active downloaders, the Infected (I) contains the users that finished with the download and they can spread the information about it through social networks, the Recovered (R) contains the users that passed through I group and after some period of time they do not spread the Video, while the Deleted (D) group includes users that removed the Video from their EHG device or the Video was automatically removed from the cache after some period of time. In the model, $N_S(t)$, $N_A(t)$, $N_I(t)$, $N_R(t)$, $N_D(t)$, $t \geq 0$, are stochastic processes representing the time evolution of each population. Suppose there are n clients of the VoD provider, they all belong to the group S at time=0, when a VoD is initially uploaded by the provider. The transition rate from state S to A consists of the probability to have a new spontaneous viewer, plus the probability to have some users that learned about the video from their social contacts and they came to a decision to watch the video. The rate can be expressed as:

$$\pi(t) = \gamma + \beta . N_S(t) . N_I(t) \tag{2.1}$$

where β is the social network contact rate for users based on the specific video, and γ is the number of spontaneous viewers that in some cases can be considerable important since a specific VoD may be advertised and promoted by the VoD provider. The transition from Active state to Infected state can be described as a poison process with mean time the duration of the video. The transitions from Infected to Recovered and from Recovered to Deleted, are considered again as poison processes with mean times estimations of the period that users are spreading the information to their social networks and about of the duration that the Video stays in the EHG device of each user. The transitions are presented in equations 2.2-2.4.

$$\delta(t) = \frac{1}{videoDuration} \tag{2.2}$$

$$\kappa(t) = \frac{1}{spreadPeriod} \tag{2.3}$$

$$\lambda(t) = \frac{1}{keepInCacheDuration} \tag{2.4}$$

The model makes the assumption that a user that downloads a video will never request it again and there are no changes in subscriber's population. The assumptions help on making the analysis simpler without losing the generality, since they are well fitting a VoD provider use case. The following equations describe the model:

$$\frac{dS}{dt} = -(\beta . I + \gamma) . S \tag{3.1}$$

$$\frac{dA}{dt} = (\beta . I + \gamma) . S . p - \delta . A \tag{3.2}$$

$$\frac{dI}{dt} = \delta . A - \kappa . I \tag{3.3}$$

$$\frac{dR}{dt} = \kappa . I - \lambda . R \tag{3.4}$$

$$\frac{dD}{dt} = \lambda . R + (\beta . I + \gamma) . S . (1 - p) \tag{3.5}$$

$$S+A+I+R+D=1 \tag{3.6}$$

Fig. 4 presents simulation results executed in Matlab when the rates are as follows: β=0.7, δ=0.1, κ=0.01, γ=0.005, p=0.7. The whole population belongs to the S state (S=1) when t=0. At time=25, the population of Active(A) increases showing that there are active users downloading the Video, while at time=50 the population of Infected(I) increases significantly. The most important line is the A(t) since it depicts the bandwidth need for covering the needs of the active downloads. It is clear that in case of epidemic spread of a specific Video, the population of simultaneous downloads is significantly increased, something that increases the difficulty in delivering high Quality of Service. A solution to the problem is the use of P2P delivery complementary to the Cloud and CDN delivery. An important observation is that by the time when the Active (A) users introduce a significant increase in their population, there is always an important number of users in Infected(I) and Recovered(R) states that can seed the Video for the them through P2P delivery method. Finally, it is clear that the transmission from Recovered(R) state to Deleted(D) does not affect the needed recovered population when it is most needed.

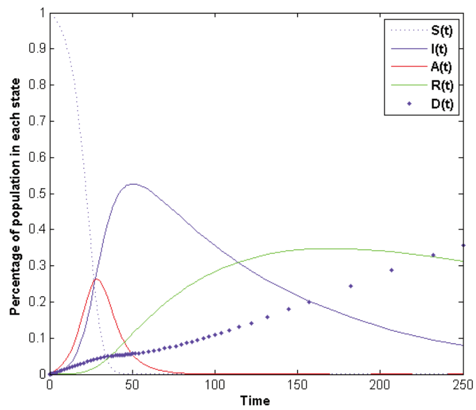

Fig. 4. Percentage of population in each State

Fig. 5 presents simulation results when the rates are as follows: β=0.3, δ=0.1, κ=0.01, γ=0.005 and p=0.2. The difference is obvious since the Active(A) population remains very low through the whole delivery process. The delivery process in this case requires fewer resources and can be possibly handled by the Cloud and CDN Servers.

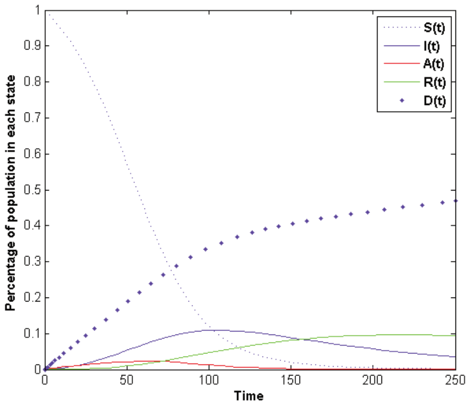

Fig. 5. Simulation Results without epidemic spread

The aforementioned epidemic model is used as a part of the Media Traffic Forecast and in the case the active downloaders are increasing rapidly the RPE can forecast the epidemic spread and take the proper management decisions for the delivery of requested media. It may be streamed: 1) directly from the Cloud, 2) through deployed surrogate servers of the CDN, 3) by establishing a Media Home Gateway Cloud (MHGC) ad-hoc system and using a combined P2P-based technology of distribution, or 4) through a combination of the above three delivery methods.

IV. Performance Evaluation Analysis, Experimental Results And Discussion

This section demonstrates the ability of the epidemic models to describe the spread of content over some content delivery systems, and the effectiveness of the Resource Prediction Engine to predict future values of network metrics of the utilization of network paths. The effectiveness of the whole proposed system is demonstrated by simulations of the usage scenarios. For the evaluation of the forecast algorithms, the monitoring data of the bandwidth usage for serving the need of a specific VoD, was collected from a VoD platform [17]. The collected measurements were for a total of 30 minutes with a period of 10 seconds, but to avoid periodicity of data, the mean value per minute was used. The 80% (24 minutes) of the data was exploited to feed the prediction engine, while the rest of the data was used for the evaluation, through a comparison between the predicted and the actual value as shown in Fig. 6. The important part of the graph is after the 24 first minutes, where it is clearly depicted that the measured values remain very close to the predicted ones.

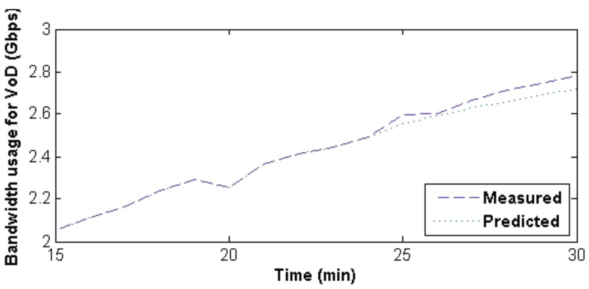

Fig. 6. Measured vs Predicted Value for the Bandwidth Usage for VoD

The prediction performance of the engine for longer term is presented in Fig.7. It includes the predicted value showing also the limit of 95% confidence and the corresponding (after the time passes) measured values of bandwidth needs for a VoD channel. The test scenarios presented are for 10, 60 and 120 minutes prediction. It is clear that the predicted values are near the actual values measured and in all cases the upper and lower limits of the 95% confidence interval include the measured value.

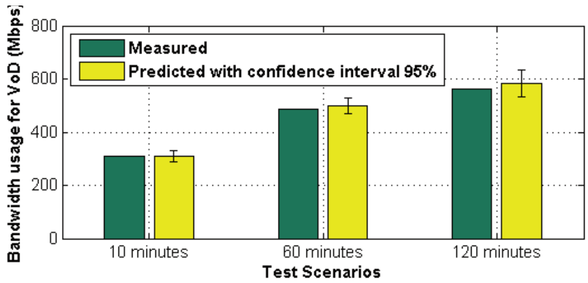

Fig. 7. Test scenarios for Bandwidth Usage of VoD prediction

Comparative performance evaluation results were extracted, to validate the reliability degree and the streaming ability of the proposed architecture, with respect to the performance efficiency. Towards implementing such scenario, a common look-up application service for video streaming is set in each node (both static and mobile), to enable nodes requesting a stream from a certain user. Fig. 8 shows the respective Complementary Cumulative Distribution Function (CCDF) that represents the sharing reliability with the download time for requests up to 20 MB. It is true that by using the proposed approach in the presence of *Reyleigh fading* and mean noise of 4dB, the reliability is not importantly affected.

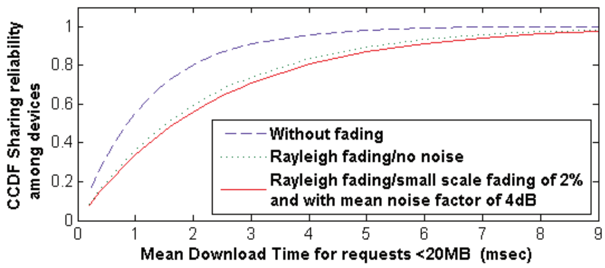

Fig. 8. CCDF sharing reliability among device with download time (msec) for requests <20 MB

V. Conclusions

This paper presents a novel network architecture that utilizes a Resource Prediction Engine (RPE), able to efficiently predict future values for network and resource usage metrics, and performs the necessary adaptation to keep a high QoE for the end users in the provision of multimedia services. An important part of the RPE is the models that describe the resource usage. This paper focuses mainly on a proposed epidemic model that takes into account the social interactions between users and manages to forecast the upcoming requests on specific content. Based on the forecast and through the management planes of the proposed architecture the selection of the optimal delivery method becomes possible. Future directions in our on-going research encompass the further study of the epidemic model for the export of multiple predicted metrics.

ACKNOWLEDGMENT

The work presented in this paper is co-funded by the European Union, Eurostars Programme, under the project 8111, DELTA "Network-Aware Delivery Clouds for User Centric Media Events".

REFERENCES

[1] Consumer internet traffic 2012-2018. CNI, CISCO. [Online]. Available: http://www.cisco.com/web/solutions/sp/vni/vni_forecast_highlights

[2] A. Passarella, "A survey on content-centric technologies for the current internet: Cdn and p2p solutions," *Computer Communications*, vol. 35, no. 1, pp. 1–32, 2012.

[3] D. Xu, S. S. Kulkarni, C. Rosenberg, and H.-K. Chai, "Analysis of a cdn–p2p hybrid architecture for cost-effective streaming media distribution," *Multimedia Systems*, vol. 11, no. 4, pp. 383–399, 2006.

[4] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, and B. Li, "Design and deployment of a hybrid cdn-p2p system for live video streaming: experiences with livesky," in *Proceedings of the 17th ACM international conference on Multimedia*. ACM, 2009, pp. 25–34.

[5] D. Niu, H. Xu, B. Li, and S. Zhao, "Quality-assured cloud bandwidth auto-scaling for video-on-demand applications," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 460–468.

[6] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. John Wiley & Sons, 2013.

[7] B. Klasen, "Efficient content distribution in social-aware hybrid networks," *Journal of Computational Science*, vol. 4, no. 4, pp. 209–218, 2013.

[8] P. Gonçalves, R. Shubhabrata, T. Begin, and P. Loiseau, "Dynamic resource management in clouds: A probabilistic approach," *IEICE transactions on communications*, vol. 95, no. 8, pp. 2522–2529, 2012.

[9] S. Sendra, L. Parra, J. Lloret, and S. Khan, "Systems and algorithms for wireless sensor networks based on animal and natural behavior," *Int J Dist Sensor Netw*, 2014.

[10] S. Meloni, A. Arenas, and Y. Moreno, "Traffic-driven epidemic spreading in finite-size scale-free networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 40, pp. 16897–16902, 2009.

[11] M. J. Keeling and K. T. Eames, "Networks and epidemic models," *Journal of the Royal Society Interface*, vol. 2, no. 4, pp. 295–307, 2005.

[12] Y. Kryftis, C. X. Mavromoustakis, J. M. Batalla, G. Mastorakis, E. Pallis, and G. Skourletopoulos, "Resource usage prediction for optimal and balanced provision of multimedia services," in *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2014 IEEE 19th International Workshop on*. IEEE, 2014, pp. 255–259.

[13] V. Assimakopoulos and K. Nikolopoulos, "The theta model: a decomposition approach to forecasting," *International Journal of Forecasting*, vol. 16, no. 4, pp. 521–530, 2000.

[14] R. J. Hyndman, M. L. King, I. Pitrun, and B. Billah, "Local linear forecasts using cubic smoothing splines," *Australian & New Zealand Journal of Statistics*, vol. 47, no. 1, pp. 87–99, 2005.

[15] Y. Kryftis, G. Mastorakis, C. X. Mavromoustakis, J. M. Batalla, A. Bourdena, and E. Pallis, "A resource prediction engine for efficient multimedia services provision," *Resource Management of Mobile Cloud Computing Networks and Environments*, p. 361, 2015.

[16] Y. Kryftis, C. X. Mavromoustakis, G. Mastorakis, E. Pallis, J. M. Batalla, J. Rodrigues, C. Dobre, and G. Kormezas, "Resource usage prediction algorithms for optimal selection of multimedia content delivery methods," in *ICC Proceedings IEEE (To appear)*, 2015.

[17] Cablenet. (2015). [Online]. Available: http://cablenet.com.cy/vod/