
Network Control Based on Smart Communication Paradigm

The increasing demand for customized and personalized services opened a new overall challenge to find scalable and sustainable solutions for the ever growing smart communications field, which supports different kinds of services, for a wide variety of future next-generation networks applications and cloud computing technologies. To address this challenge, new and cross-disciplinary approaches are required to optimize the distribution of media data which needs to explore how smart communications, built from user perception and feedback, affect protocols, global design, equipment, algorithms, paradigms, power consumption, etc., for a large family of applications using different network technologies. The key observation behind this chapter is that the addressed problem raises many interesting challenges when considered from a control theoretic perspective. First, the general framework is presented. Then, the main inter-disciplinary views are described in detail.

1.1. Motivation

The last two last decades of the 20th Century were driven by the emergence and evolution of the Internet and its use for human communications. There is no doubt that in the 21st Century, the concept of Internet of Things (IoT) and its applications will have a key role in the way we understand our society. The IoT describes the trend for environments, buildings, vehicles, clothing, portable devices and other objects to have a digital representation and the ability to sense, use or exchange information. IoT makes things more interesting by connecting real-world objects, places and people through the digital world. Small objects connected together as part of the IoT today are considered one of the main challenges and the business revolution for the coming years. A widespread use of such connected objects will influence people, societies and businesses. IoT will transform how we live in our cities, how we travel, how

we manage our lives sustainably, how we age and how we engage with services and entertainment.

Over the years, the continuous technological evolution and the development of new applications and services have steered networking research toward new problems, which have emerged as the network evolves with new features towards what is usually referred to as the future internet, which has become one of the basic infrastructures that supports the world economy nowadays. In fact, there is a strong need to build a new network scenario, where networked computer devices are proliferating rapidly, supporting new types of services, usages and applications: from wireless sensor networks and new optical network technologies to cloud computing, high-end mobile devices supporting high-definition (HD) media, high-performance computers, peer-to-peer networks and various platforms and applications. This new network scenario is fueling research in the area of new network architectures that consider both the requirements and demands of key emerging applications and services, such as cloud computing and Internet video, and the currently deployed network infrastructures.

Conversely, media distribution over the networks is now growing at a pace that will very soon threaten to change networks to a degree at which the perceived quality will no longer be acceptable. Also, users are switching from operator-controlled IPTV services to uncontrolled over-the-top media applications. High-quality delivery platforms increase the user expectation of quality, equivalent to the quality of traditional broadcast distribution. Furthermore, live TV, for example during sports events, may have a large impact on the amount of traffic in small geographic areas. Various distribution formats, delivery platforms and methods for protecting material leads to many copies of the same material that must be produced and distributed in parallel. Transcoding of content to different formats requires a lot of energy for both calculation and cooling, which is at odds with modern standards of energy efficiency and long-term durability. Also, users become connected to the Internet through different devices (mobile, PCs, HD displays, etc.) and networks (3G/4G/UMTS, ADSL, VDSL, etc.). Devices and networks can vary a lot in their features and capabilities. Hence, each user has a specific context. This situation produces a very heterogeneous environment where an increasing demand for customized and personalized services exists.

Therefore, the overall challenge is to find scalable and sustainable solutions for the ever growing smart communications field, which supports different kinds of services, for a wide variety of future next-generation network applications and cloud computing technologies. To address this challenge, new and cross-disciplinary approaches are required to optimize the distribution of media data. One approach is to introduce and combine methods for more effective storage of the data close to the consumer and various types of peer-supported distribution mechanisms. Decisions on when and where data are temporarily stored should be based both on estimates of

expected demand and the final consumer demands on perceived quality. Also, statistical methods for online estimation of consumer demands will be crucial. Many of these issues are complex and can, in the short term, hardly be solved by any single approach. We need to find solutions to deliver network services in the most efficient way to provide users with the best perception while taking into consideration scarce network resources.

We also need to explore how smart communications, built from user perception and feedback, affect protocols, global design, equipment, algorithms, paradigms, power consumption, etc., for a large family of applications (healthcare, subaquatic, vehicular, robotic, economics, etc.) using different network technologies (Sensor Networks, MANET, VANET, Ubiquitous, Virtualization, Data centers, etc.). Indeed, autonomous applications embedded in complex configurations and dynamic environments have experienced a rapid expansion from classical applications where different modular devices, actuators and sensors interact closely. This expansion has considerably impacted the control of a given system in a centralized manner. Current research trends propose new autonomic architecture schemes that manage and control future emerging networks: Sky of clouds, IoT, Smart Grids, Smart Cities, etc. In parallel, the evolution of Internet usage appeals for higher quality guarantees in order to support stringent services. In this context, for example, healthcare and wellness applications such as helping elderly people, assisting dependent people or habitat monitoring in a smart environment, constitute some of the potential scenarios of convergence between autonomous systems and smart network technologies. These applications, which are based on high-level commands, accomplish some specific tasks, reveal new challenges regarding mechanic design, portability, acceptability, power support and efficiency, control theory, etc. In addition to achieving portability and low-power systems, major challenges that substantially limit the efficiency of any autonomous system must be addressed. This new vision will highlight the overlapping of two domains, namely autonomous systems and smart network technologies, devoted to different applications and a large variety of domains.

1.2. General framework

Our vision of a new paradigm is to make interactions first-class objects from the perspective of the user, the application and the network components. This is achieved by analyzing the interaction between the user and the application with quality perception metrics that are used to fix the control/command chain in network components. The idea here is how to integrate these metrics into a control/command chain in order to construct a network system. The key idea is to design an adaptive loop for improving the service quality of a network system by taking into account the end-user feedback. In this model, the perception measurement module evaluates the user feedback and sends the result to the control module. The latter analyzes the received results, makes a decision and sends it to the command module that initiates

the chosen action onto the network system. This loop model refers to an adaptive system based on the end-to-end user's perception and results in smart communications between equipment.

The key observation behind this vision is that the addressed problem raises many interesting challenges when considered from a control theoretic perspective. The characteristics of multimedia contents and the network resources vary over time. There are propagation delays in networks and the quality of offered services is still difficult to evaluate. For example, multimedia contents are strongly linked to the observed rate distortion. In fact, techniques borrowed from control theory, ranging from stability and controllability techniques, can be applied to adaptive, predictive, or stochastic control in both centralized and distributed multimedia delivery architectures. Unexplored challenges in control techniques for efficient multimedia delivery will be emphasized.

The use of artificial intelligence tools together with biologically inspired techniques, to control network behavior in real time so as to provide users with the Quality-of-Service (QoS) that they request, and to improve network robustness and resilience, is necessary. The key idea is based on control theory and machine learning techniques in order to support smart communications driven from user's perception to avoid any complete interruption in the whole chain of service treatment.

This is not the first attempt to get rid of the notion of technical control mechanisms in network components based only on network performance. The operators always try to deliver network services in the most cost-effective and resource-efficient manner with assured user perception. Therefore, service providers are switching their focus from traditional control mechanisms to user satisfaction, which is the overall success of a network from the user's perspective. According to Daniel R. Scoggin, "The only way to know how customers see your business is to look at it through their eyes". In fact, the convergence of network technologies has been driven by the converging needs of end users. The perceived end-to-end quality becomes one of the main goals required by users that must be guaranteed by the network operators and the Internet service providers, through manufacturers' equipment. This is referred to as the Quality of Experience (QoE) notion, which is commonly used to represent user perception. QoE is not a technical metric, but rather a concept consisting of all elements of a user's perception of the network services such as those found in the field of affective computing. To ensure good end-user perception, service providers have encountered QoE challenging issues, such as QoE monitoring, measurement, diagnosis and management. Therefore, all of network functions such as admission control, access network selection, routing, resource allocation, transmission control and server selection are expected to be adaptive to user QoE.

This vision differs from others by focusing on control theory and machine learning techniques in order to support smart communications to avoid any complete interruption in the whole chain of service treatment based on the user's perception. The bio-inspired model is one of the possible ways in which machine learning addresses the large-scale and distributed issues in new environments such as massively large-scale, wide-area computer networks and mobile *ad hoc* networks. These new environments also present new challenges because they are highly dynamic, sometimes unreliable, and often span over a wide area. The aim here is to demonstrate that these models based on living organisms can effectively organize themselves, in large numbers of unreliable and dynamically changing components, into structures that implement a wide range of functions. In addition, most biological structures exhibit several favorable properties such as robustness to failures of individual components, adaptability to changing conditions and dynamic scenarios and their low reliance on explicit central coordination.

Many works have been done based on continuous learning paradigm. For example, in our early work [MEL 05], we demonstrate that the approaches based on central and explicit control over the system are also not suitable for large-scale distributed networks because they are not scalable. In addition, centralized approaches introduce a single-point-of-failure, which should be avoided whenever possible. The article introduces an adaptive model that is very useful in tracking a phenomenon that evolves over time. The problem was naturally formulated as a dynamic programming problem, which, however, is too complex to be solved exactly.

The most important challenge is the use of the concept of a state-dependent model based on a global optimization cost function whose parameters are learned from the real-time state of the network with no *a priori* model and applied in several domain applications. This kind of model is based on the fact that it is capable sampling, estimating and building the network environment's model of pertinent and important aspects, such as type of traffic, QoS policies and resources. It is based on the trial/error paradigm combined with swarm adaptive approaches. In a more recent work, we developed a global system that uses a model which combines both a stochastic planned pre-navigation for the exploration phase with a deterministic approach for the backward phase.

We have studied and introduced the concept of control mechanisms driven by user perception. In fact, we observed that the most important challenge of developing an efficient QoS mechanism is to resolve the NP-completeness complexity problem in the presence of more than two non-correlated criteria [GAR 79]. The solution proposed is to integrate the user's perception to control the time-evolving optimization cost function in order to maintain a certain satisfaction degree of end users, which is the objective of all network systems, without optimizing all network technical parameters.

The development of individual perception estimator agents, including the human perception models, the service models and the subjective test methodology, and the absence of applications address the problems of smart communications outlined in section 1.1: scalability can be achieved by designing metrics that support large data sets; platform diversity is supported by the ability to use services that take advantage of the platform's interaction resources and by a natively distributed architecture; flexibility is supported by the independence between metrics and the services that they operate on; and heterogeneity is supported by interconnecting different kinds of quality estimators.

1.3. Main innovations

The following sections describe several main inter-disciplinary views in order to accomplish the main tasks of the development of this new vision: developing at the theoretical level the appropriate unified formal model based on bio-inspired modeling and knowledge distribution in order to construct a scalable and robust environment with low complexity for large-scale dynamic networks; studying at the empirical level why and how user perception can be quantified, analyzed and influenced; creating at the conceptual level a general framework for protocol stacks dedicated to smart communications featuring multiple technologies, devices and users; and building at the engineering level the appropriate model of programming abstractions and software architecture to develop a full-scale framework.

1.3.1. *User perception metrics and affective computing*

As a critical measure of the end-to-end performances at the service level from the user's perspective, the user's perception is a kind of metric for the system design and engineering processes that can only be measured dynamically at the very end of any transmission activity.

To clarify the above definition, many standards organizations have studied this concept. We cite here three relevant standardization organizations:

- ITU-T: Study Group 12 is the main group studying the QoS and QoE in the ITU Telecommunication Sector (ITU-T). Some important recommendations include: ITU-T G.1010 [REC 01a], ITU-T G.1030 [REC 05, REC 11b] and ITU-T G.1070 [REC 07].

- Broadband-Forum [BRO 95]: the user's perception is considered as the measure of overall performance from the user's perspective.

- TeleManagement-Forum (TM Forum) [TMF 88]: TM Forum considers a set of key quality indicators (KQIs) and key performance indicators (KPIs) as measurements of perceived quality rather than network performance.

In fact, even though user perception is difficult to predict due to its subjective nature, defining a set of metrics is necessary for the assessment of overall service quality. One way to achieve this is using the top-down approach in which the factors that influence user perception are determined at the beginning. Then, this information is used to generate operating requirements. Finally, a method to constantly evaluate these factors and improve them is created based on reference measurements (absolute and/or relative evaluations), preservice and in-service tests (achieved offline or online), and problem detection (image quality, audio quality, video quality, interruption, task completion, time, etc.). Also, the perception control procedure, i.e. mapping of the user's perception onto the network performance, is an important challenge. To achieve this goal, we have to identify the relationship between network performance metrics and their effects on user perception [ARO 12].

Also, a critical step in managing and evolving the network and business is to identify and measure the right KQIs. When identifying KQIs, there will be as many different expectations as there are users, but most of these expectations can be grouped under main categories to be defined. For example, in network services, the reliability dimension has some KQIs as follows: service availability (anywhere), service accessibility (anytime), service access time (service setup time) and continuity of service connection (service retainability). On the other hand, for telecom applications, KQIs could be as follows: connection establishment success rate, connection drop rate, handover success rate, throughput and block error rate. In fact, the KQIs are measured for each service in order to calculate an estimate of each service score.

The other crucial question here is to develop systems and devices that can recognize, interpret, process and simulate human affects. It is an interdisciplinary field spanning computer sciences, psychology and cognitive science [PIC 97]. Detecting emotional information begins with passive sensors that capture data about the user's physical state or behavior without interpreting the input. The data gathered are analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expressions, body posture and gestures, while a microphone might capture speech. Other sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance. Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done using machine learning techniques that process different moralities using speech recognition, natural language processing or facial expression detection, and produce either labels (i.e. "confused") or coordinates in a valence-arousal space.

Conversely, different sources of uncertainty and imprecision may arise with these collected measurements. Such imperfections may be due to the imprecision of many environmental aspects: signal, data link, network, etc. For example, it may be due to communication links that are unreliable, either due to operational tolerance levels or environmental factors. The theory of belief functions can be used here to represent, correct and reason with partial information [SHA 13].

1.3.2. Knowledge dissemination

As explained earlier, the objective here is to transform the current network centric view to a future user-centered view. To make appropriate decisions, such a network needs to collect a quantity of knowledge, which forms a new knowledge plane. The knowledge plane must be created autonomously, continuously and dynamically. The knowledge must be acquired in both a vertical manner (knowledge emanating from different layers of the same node) and in a horizontal manner (knowledge coming from multiple nodes). However, the most important factor is to have an entire distributed knowledge plane [LI 08]. To do this, knowledge must be stored (physically and logically) in different parts of the network.

This knowledge plane must meet three main features [SOU 12]:

- The knowledge must be collected in real time using various monitoring tools on different network layers.
- An analysis and learning component can make the best decisions based on the context.
- The knowledge must be shared among different nodes.

It is clear that the knowledge dissemination in the network remains an open problem. In fact, previously realized works on the autonomic paradigm are mostly based on a multi-agent. An agent is a computer system located in an environment and acting autonomously to achieve the objectives for which it was designed. By distributing agents on each equipment, the system can handle local problems locally, based on all acquired and taught knowledge [KEP 03]. In most situations, local problems are more easily resolvable than global problems, especially as a local problem is detected much faster than in an approach called centralized. Obviously, this type of architecture is not realistic because the knowledge dissemination will generate a very high overhead.

From a different point of view, many service quality parameters influence end-user satisfaction. Hence, designing Service-level agreement (SLA) oriented resource management schemes are required for service operators that provide attention to each customer. For instance, this can be achieved by enabling communication to keep end users informed and obtain feedback from them, increasing access and approachability for end users and understanding the specific needs of each end-user flow.

Moreover, service requirements of users can change over time and thus may require updating of original service requests. For example, the resource management algorithms will be able to self-manage the reservation process continuously by monitoring current service requests, predicting future service requests, and adjusting schedules and prices for the current communication.

Hence, self-configuring components to satisfy new service requirements are needed, so that more autonomic and intelligent providers can better manage the limited supply of resources with dynamically changing service demand. For end users, the system acts on their behalf to select suitable providers and negotiate with them to achieve ideal service contracts. Thus, providers also require autonomic resource management to selectively choose appropriate requests to accept and execute depending on various operating factors, such as the expected availability and demand of services (both current and future), and existing service obligations.

1.3.3. *Bio-inspired approaches and control theory*

Many studies have been conducted over the last decade for an alternative paradigm that would address the integration of dynamic criteria evolving in time with uncertainty. The most popular formulation of the optimal distributed optimization problem in a network is based on a multiparameter flow optimization whereby a separate objective function is minimized with respect to the types of flows subjected to multiparameter flow constraints. However, due to their complexity, high processing overheads, a few proposed schemes could be accepted in real life. For a network node to be able to make an optimal control decision, according to relevant performance criteria, it requires not only up-to-date and complete knowledge of the state of the entire network but also an accurate prediction of the network dynamics during propagation of the message through the network. This, however, is impossible unless the control algorithm is capable of adapting to network state changes in almost real time. So, it is necessary to develop a new intelligent and adaptive control algorithm. This problem is naturally formulated as a dynamic programming problem, which, however, is too complex to be solved exactly. In these adaptive paradigms, the environment is modeled as stochastic (especially links, link costs, traffic, congestion, etc.), hence control algorithms can take into account the dynamics of the network. However, no model of dynamics is assumed to be given. This means that these approaches have to sample, estimate and perhaps build models of pertinent aspects of the environment in order to support a framework that addresses the identified trade-offs needed for an overall control mechanism.

Many different approaches based on bio-inspired paradigms, such as cognitive packet networks [GEL 04], ant colony optimization and reinforcement learning paradigm [SUT 98, MEL 11a], exist in the literature, and the objective is to see how to reduce their complexity. In addition, such a network system must also design and implement intelligent and adaptive optimizing control algorithms to support irregular and differentiated flows, and take into account the network state and its dynamics. The use of these bio-inspired models is one of the possible ways to address the large-scale and distributed issues [DRE 10]. It is well known that living organisms can effectively organize themselves in large numbers of unreliable and dynamically changing components into structures that implement a wide range of functions. In

addition, most biological structures exhibit several favorable properties such as robustness to failures of individual components, adaptability to changing conditions and dynamic scenarios, and their low reliance on explicit central coordination. Currently, the main research direction in the bio-inspired modeling area focuses on the exploration of solutions that enable the construction of a scalable and robust environment with low complexity for large-scale dynamic networks. Such an objective can be met not only by relying on the decentralization of functions or their self-organization but also by depending on the methods, techniques and tools for data retrieval and analysis and learning, all of which participate in the global decision process (thereby simplifying and minimizing human actions and interventions).

1.4. Conclusion

The addressed problem raises many interesting challenges when considered from a control theoretic perspective. The characteristics of real-time service contents are time-varying as are the network resources. Control theory offers interesting solutions for several problems related to the control of networks, such as the stability of protocols in order to guarantee a level of performance on all the space of uncertainties. This new vision will focus on techniques borrowed from control theory, ranging from stability and controllability techniques, to adaptive, predictive or stochastic control in both centralized and distributed network delivery architectures.