Deep Science Publishing
https://doi.org/10.70593/978-81-989050-5-5_9

Chapter 9: Self-configuring networks and embedded intelligence in autonomous communication systems

9.1. Introduction

Recent advances in micro-electronics have catalyzed a new wave of progress in electronic and electro-optical as well as wireless and opto-electrical circuits by enabling the convergence of computing, communication, and control technologies into intelligent systems, or intelligent devices. Affordable and aesthetically integrated, these intelligent devices will soon proliferate within commercial buildings, homes, cars, and clothing, as well as at factories performing manufacturing, automation, and logistics. These embedded intelligent systems will offer unprecedented levels of convenience, quality of life, safety, personalized services, and productivity enhancements. Self-configuring networks for communication and computing between various embedded intelligent systems will thus become a “Sixth Utility” in addition to power, water, sewage, gas, and telecommunication (Chen et al., 2016; Han et al., 2016; Li et al., 2020).

In the near future, we envision millions of embedded intelligent systems with wireless short-range communications for sensing, monitoring, tracking, controlling, or managing various functions, such as building management and automation, home security and emergency response, food and drug storage and delivery, vehicle assessment and fleet management, and clinical monitoring and healthcare. For localized themes, users can plug in a sensor watch, e-mail a friend, and look for fun spots for the day without tedious plans. Users in the hotel can book the check-out room time, and mail to the airline for the flight details. With the tremendous reduction of cost and size, user-unfriendly specifications, like address, link, and key management, are also no longer acceptable for laying down pervasive and ubiquitous communication networks. Instead, as wireless communication systems become part of the intelligent systems for various applications, self configuring and self-optimizing capabilities are imperative. Additionally, the popularity of the Internet has sparked the advent of increasingly autonomous communication systems. Penetrating devices able to communicate and access information over the Internet from any location will soon be commonplace. This ubiquitous communication capability and revolutionizing advancement is accelerating the creation of a communication-oriented society (Sze et al., 2017; Zhang et al., 2019; Li et al., 2020).

Fig 9.1: Self-Configuring Networks

9.1.1. Background and Significance

The experimental demonstration reported here uses an impulsively launched network, which builds itself to support a mobile electronic speech relay over a specified time. The network employs two novel techniques: lightweight self-organization from arbitrary locations and sources, and embedding of a selected backbone message relay function in itself. Service interruption and the time of network configuration are minimized. The speech delay through the radio relay is not noticeable. Quality, voice activity detection, and the bi-directionality of speech conveyance are reported competent by listeners. The employed semi-automatic network configuration and embedded relaying system concepts are general, apply to any wireless technique, including terrestrial operation, and can be used by existing and future radio systems. The presently demonstrated concept is extensible to higher numbers of nodes, higher data rates, local calibration, improved static and dynamic pattern rejection, inter-user interference suppression, and to larger search-cover volumes.

Since the inventive approach allows for arbitrary spatial distribution of the relay and terminal radios, addresses an extremely low-power relay mode of operation, the established time delay through the terminal radio in the full-duplex mode is low, and it can work with small and cheap mobile radios, electronic speech relay represents an attractive solution for many applications. There is no electromagnetic homing device to unmask eavesdropping relayers during and after message conveyance. The possibility of sampling or replaying message content stored in the terminal relay radio is eliminated. The relay surrogate can be made inobtrusive. The timing is completely determined by the terminals. No menuing is required.

9.2. Overview of Autonomous Communication Systems

Today, the area of autonomous communication systems is at the intersection between many research topics throughout next-generation telecommunications research and development. We all want significantly increased capacity for all types of communications – wired and wireless, mobile and stationary. We need a user experience that is intelligent and autonomous, without having to bother to understand or manage any technicalities. We simply want it to work, and to operate on our selected terms. We expect wireless communication to work at the same levels of reliability, availability, and quality of service that we expect from wired systems, particularly for integrated information-centric services – which is a major area of research clustering today. We want to be able to integrate our autonomous intelligent systems with our autonomous intelligent vehicles, creating a global network of interacting autonomous networked systems – an autonomous sensor grid, or whatever term happens to be fashionable when the systems are operational. Finally – and this is really the main motivation for increased autonomy – we want to able to off-load menial tasks to our intelligent systems, in support of our own activities, leaving us free to do what men do best, because we can, and because there are tasks that are better performed by men than by machines, leaving the menial tasks to the machines. The autonomous and self-configuring communication systems area encompasses many forms of autonomous distributed systems that show similar types of intelligent system behavior – they work for us, rather than we working for them. Communication forms a key subgroup of these intelligent distributed systems activities. Autonomous behaviors manifest themselves from pre-configured systems, whether wired, wireless, mobile, or otherwise non-static in deployments, either temporarily or permanently dedicated to specific applications. Security is an inherent characteristic of any intelligent distributed system.

9.2.1. Definition and Scope

Autonomous communication systems (ACS) are a new type of communication system. Their fundamental characteristic is that the fundamental functions of information exchange, protocol negotiation and operation, resource and service management, fault management, and performance assessment and adaptation are performed in a largely autonomous or semi-autonomous manner with little to no human intrusion. This means that a large part of the lifetime of the system is spent in a self-configuring, self-defining, and self-operating operational mode, independently of detailed support from human engineers or operators. Although the manifestations of this newly achieved independence might differ within different types of systems and environments and during different phases of the system lifetime, this technological evolution endows communication systems with a flexible intelligence based on factors like a tight coupling of hardware and software with real-time adaptation to changing operating conditions, embedded and distributed kernel intelligence, adaptive, application-aware, deeply layered and intrasystemally heterogeneous protocols, high levels of reliable automation, reliability through adaptation and redundancy, enhanced performance visibility and assessment capabilities, and highly modular and open architectures.

The use of technology, systems, protocols, and tools that were generally referred to as embedded and distributed technology, which gave rise to the new classes of what relatively recently have been conceptualized as E&D systems whose main distinguishing characteristics were the high levels of density, low costs, and built-in intelligence they offered, is the basis for the self-configuring nature of ACS. The primordial concept, which defines an ACS and frames their main goals and objectives, was the ability to configure themselves in an autonomous, spontaneous manner in response to disastrous perturbations and failures. They were thus obtained from the successful combination of two fundamental E&D technological concepts, which were also the main enablers of their low costs, the possibility of integrating communication and computing functionality at ever-lower costs into small modules, and the availability of communication and computational infrastructures that were progressively omnipresent.

9.2.2. Historical Development

The increasing complexity of communication systems due to the introduction of new functionalities and services, the increase of users and their mobility, and the increase of frequencies of terminals have modified the traditional concept of networks and made communication systems truly complex systems, where communication nodes cannot be considered as mere signal forwarding, storage, and regeneration units. Instead, they are becoming intelligent nodes that are capable of autonomous decision-making. The bottom-up concept of embedded intelligence in communication nodes has been gradually applied through the introduction of advanced self-managing capabilities in autonomous communication systems, such as self-configuration, self-healing, self-optimization, self-protection, and self-assurance.

While decentralized or self-organizing wireless communication systems for short-range communication applications have already been introduced as commercial products, the concept of embedded intelligence for self-managing functionalities is progressively finding application in advanced wired and cellular mobile communication systems. Indeed, the concept of embedded intelligence is fundamental to achieve the required flexibility and efficiency in future communication systems, characterized by the introduction of new services and the change in offered quality of service in terms of delay, throughput, and security, sometimes within very short periods of time.

The vision of communication networks and systems able to autonomously prepare and configure in order to provide optimized services, following the dynamic changes of external conditions and performance parameters, has been studied in numerous research works and is reflected in numerous ongoing projects worldwide. From a formal viewpoint, the concept of autonomous communication systems dates back to the idea of autonomic communication.

9.3. Self-Configuring Networks

Self-configuring networks are dynamic self-organizing and self-managing networks for an autonomous communication system or network that in addition to embedded intelligence provides mechanisms for self-discovery and service discovery and is steerable using a collaborative multi-agent-based system. A self-configuring network does not require any specialized user knowledge in order to set up and deploy as a blocking factor for offering services without the need for a highly skilled workforce. To address this need, self-configuring multi-hop networks are proposed where the network connection function optimizes and autonomously drives the deployment of the system at deployment time and the configuration and orchestration of the system at service provision time. A mesh of sporadic and multi-hop self-configuring networks where each multi-hop network is built on demand with a point-to-point link to cover a specific communication delay budget with dynamic adaptation properties to reroute flows if the delay budget over the link is exceeded while the source and intended destination system nodes are periodically self-discovered using a common self-discovery application software agent acting on the network itself is proposed. After deployment, the nodes of the multi-hop network are again periodically discovered by another peer self-discovery application agent that implements the application-layer logic for the actual application service provision.

The purpose is to configure service networks end to end. At the intended service destination, the intuitive interface of application services evokes a self-discovery agent in parallel to the delay budget agent. The role of the intended destination service is to apply for the allocation of the requested resources and the ability to offer such a service at such a cost. The end-to-end configuration of a specific network determines the multi-hop delay budget requested and at what cost by multi-point collaboration, and the connection function predicts which configuration is suited or not using application-layer preferences driven by reasoned temporal knowledge.

9.3.1. Concept and Architecture

Transitioning from conventional to self-configuring networks, the successful deployment of future autonomous communication systems relies on distributed intelligent control mechanisms. In self-configuring networks, no unique or centralized authority is available for installation, configuration, and management of the system. Self-configuring refers to full decentralization and autonomy. Distributed intelligence ensures the convergence and stability of system operation despite the absence of inbuilt authority. With these properties, embedded local intelligence is inherent in the building blocks of the network – nodes, gateways, and mobile users. Building blocks might include slightly intelligent nano- and microsensors, switches, routers, and gateways, exchange information about themselves and their environment and respond autonomously for optimal and efficient operation of the entire self-configuring network.

Self-configuration is a two-layer distributed intelligent control architecture. The lower layer allows building blocks to exchange information and take in timely actions about critical operational variables to ensure parameter stability, determined based on knowledge on basic frameworks and physical laws governing the special communication functions regarding the application area. The upper layer allows selection of safety and efficiency criteria, during both normal and failure conditions, and determination of response rules. In a self-configuring system, performance- and risk-related criteria and rule sets are determined based on adaptive and self-learned parameter stability knowledge within available resources, as required for local resolution of conflicts with those determined at the upper layer. The lower layer controls operational parameters in real-time, implementing the response rules from the upper layer, possibly adjusting the upper layer output. This architecture provides for enabling decentralization and autonomy in a fully intelligent mode.

9.3.2. Advantages and Challenges

The technological capabilities of integration, miniaturization, and consolidation of both hardware and software functions of embedded systems generate a very significant impact in many areas of technological developments. In the networking area, wireless self-configuration without any previous configuration or any long term human intervention not only permits the deployment of a large amount of miniature tiny systems for all purposes but also creates a unique condition for the systems to become autonomous, having their own intelligence and control. Self-configuration networking becomes an enabling technology for the realization of autonomous or semi-autonomous objects or units having embedded wireless connectivity, that is, self-configured embedded intelligent communication systems, such as Wireless Sensor Networks, Advanced Embedded Systems, Internet of Things, etc. The technological trend leads to embedded systems built with very light-weight hardware and software architecture at low cost, providing simplified functions and services requiring very low power.

Fig 9.2: Advantages and Challenges of Self-Configuring Networks

In such a scenario, traditional network functions such as aiding the self-configuration and sensor activation, specification and management of policies and operational conditions, supervision of data communication flows and error handling are replaced by much lighter programs or stripped embedded systems. Despite the network providing only basic functions, such as self-configuration and dealing with routing and error messages, the intention is to generate an architecture so simplistic that all systems or entities may become totally autonomous and with very low requirements in terms of power and communication processing. In addition, given the extensive logarithmic growth of energy density in wireless networks, the amount of time during which it is possible to operate an embedded system without battery charging or battery replacement is expected to be very large.

9.3.3. Use Cases in Industry

Industry is finally starting to realize the full potential of self-configuration, and various private and government initiatives have been developed to explore some of the possibilities in this area. The concept of self-configuration goes far beyond the stand-alone self-configuration capabilities offered in most of the widely deployed wireless area standards. These solutions are typically limited to ad hoc autonomous startup, and the configuration is limited to configuration options needed for startup purposes, like network startup and maintenance. However, the possibilities of self-configuration are much broader and entail self-configuration also during the operational stages, or while elements are active and carrying traffic.

The most ambitious visions are illustrated in some of the research funded by government initiatives. The Adaptive Mobile Wireless Network initiative is exploring. The goal of this program is to develop wireless communication systems that are capable of self-configuring to temporarily meet rapidly diverging communication needs and new access capabilities. Under this program, systems are being developed which in addition to the stand-alone configuration capabilities have global optimization potentials. These systems are designed to offer wireless converged services to both military and commercial users. Other research involves low-Earth orbit satellite constellations for both communications and remote sensing. The basic idea is that a network will automatically adapt its operation to the need of both specific services and overall network throughput and delay.

9.4. Embedded Intelligence

Embedded intelligence has brought about unique capabilities that help network devices to gather, process, and generate the key information within their operational environment. This allows for the generation of a data-driven view of the operational context where the network device operates, enabling informed local decisions and actions to be taken in a timely manner and independent of a central system. To produce meaningful data-driven situational awareness output, the embedded intelligence system residing in the network device should possess sufficiently high levels of autonomy. This implies an ability to operate through long periods without the need of external resources for its operation.

Role of Artificial Intelligence

Artificial intelligence plays an essential role in self-contained operation of the embedded intelligence system. This is because operating without external resources implies that the embedded intelligence system will only have access to the low-level sensory measurements available from the sensors embedded in the network device for its intelligence processing. Such low-level sensory data lacks any form of redundancy and therefore makes it difficult to perform the different and diverse data processing tasks involved in embedded intelligence processing. What is needed is an AI capability that is lightweight enough that it can be embedded in the resource constrained embodiments of the network devices while being sufficiently rich to enable the flexibility needed to operate in the variety of operational environments encountered.

Machine Learning Techniques

Machine learning has emerged as the preferred approach to creating the much needed embedded AI capabilities. Some of the unique aspects of the learning capability required for embedded intelligence processing is the small data regime and the lack of training data. Embedding intelligence in a product for which sufficient training data exists is hard enough – the problem becomes even more difficult when such training data is absent.

Data Processing and Analytics

The majority of AI-based embedded intelligence systems rely on a certain degree of input structure. There are numerous methods for achieving this, but the rule here is that no learning can be performed for input-output pairs that hold no structure. Unfortunately, the vast majority of embedded intelligence applications have no such internal models. Possible solutions for this limitation include hybrid systems with an external model.

9.4.1. Role of Artificial Intelligence

Despite being the main theme of this book, self-configuration did not – and probably could not – rely uniquely on intelligent techniques for understanding and adapting autonomously over time to the ever-evolving user preferences and operational contexts of wireless communication provisioning and manufacturing processes. The term intelligence, in fact, could be understood in the sense that network entities are able to evaluate previous or ongoing events to adapt or configure themselves towards future situations leading to a given desirable utility function. In this context, Artificial Intelligence, and especially its machine learning branch, is becoming an enabler and a facilitator for the self-configuration of network entities. In the next few years, communications service providers will be asked to innovate, optimize, secure, and automatize their services, while minimizing costs. Self-configuration of network processes is indeed a critical issue to allow service providers to keep costs down and automate numerous manual operations, but also to better design their service portfolios. Depending on the profiles and expectations of different kinds of users, and on the operational context, service providers should design their portfolios of wireless services, and match demand and offer, to enhance user experiences and revenues while minimizing costs. Machine learning will help service providers design portfolios for specific user profiles and operational contexts. In a further stage, intelligence would enhance services satisfying specific user demands, and network and service performances, during the entire operational life. Anomaly detection in network and service processes, user-driven recommendation of different usage modes, or of alternative services to select in specific operational contexts would all be functions of learning. Supervised learning from experience is easy to understand for users and for designers: the quality of network-driven recommendations is based on the number of experiences and the accuracy of previous recommendations. Generative AI could also develop additional service functionality.

9.4.2. Machine Learning Techniques

Machine Learning is a tool that enables the development of systems that can extract knowledge from data. A significant amount of stored data is often available in Communication Systems. This data can provide useful insights about system operation and behavior. Thus, the application of Machine Learning comes natural to Communication Systems. The decision system has to handle vast amounts of data in a near real-time manner. This is one of the important challenges in the deployment of Machine Learning in an Autonomous Communication System. Another challenge relates to the fact that the data often comes from devices with constrained resources. The applications of Machine Learning techniques in Communication Systems has received significant attention in the last years.

The informal definition of Machine Learning is “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” The above definition encloses many aspects of Artificial Intelligence: It encompasses work on unsupervised and supervised systems, systems using symbolic and sub-symbolic representations, as well as systems where the learned knowledge has limited applicability and systems capable of acquiring and exploiting deep, domain-specific knowledge. In Computer Science, “Machine Learning is a field of study that gives computers the ability to learn without being explicitly programmed.” The above definition comprises systems capable of refining a decision performance after performing a task.

The Machine Learning methods are often grouped into three categories: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the learning algorithm is provided with labeled training data, and the task of the algorithm is to learn a function that maps inputs to desired outputs. Training data is usually labeled manually.

9.4.3. Data Processing and Analytics

Communications networks and systems constituting Internet infrastructure generate massive amounts of data. This data, most of which is unstructured and comes from different sources, is typically stored and never again touched. We propose that communication networks and systems use this data to create value for communication services and users and for the networks themselves. For example, communication services and user experience can be enhanced by analyzing logs of service usage and surveillance data collected by the providers of communication systems and by communications service providers.

Additional value can be provided to communications service providers, as well as to the autonomous systems constituting the networks of Autonomous Communications, by machine learning as well as advanced data processing and analytics applied to Big Data generated by the systems and networks themselves. For example, operations of communications systems, their environments, and services offered can be more reliable through predictive analytics of data collected from systems that are implementing cyber-physical systems infrastructures such as Smart Cities Communications, Smart Grid Communications, Smart Home Networks, and Smart Automotive Networks. To this purpose, Artificial Intelligence tools should be developed that meet the specific requirements of Autonomous Communications. Broadly, communications system, network, and service infrastructure requirements are created by the urgency for performance reliability, availability, trustability, predictability, and quality-of-service provisioning in cyber-physical and in Autonomous Communications Infrastructures.

9.5. Communication Protocols

Communication protocols have been elaborated to provide a sufficient level of support for the communication nodes in heterogeneous distributed object-oriented systems, which may contain service and application servers, gateways and clients (including mobile clients). Protocols allow implementation of secure and reliable end-to-end communications. Communication legislation for secured information services is specified and previously predefined for all nodes, and policy files for each player's communications are updated in on-line guidance. Binaries generating output for selected device classes allow on-line optimization of the generated code for remote devices, and download content with variable granularity. Semantic communication during assisting requires certain network communication services not provided by the current protocols, or requiring the protocol adaptation. These services include: low- and high-priority, session establish manager, enabling distributed multi-level prioritizing of all communication flows in the network, noise reduction manager and digitally animated remote user's avatar manager.

We also need to define a mechanism for self-organization of topology relations as a reward or punishment for the discovered activity, through establishment of temporary devoted dialog channels. Development of sophisticated application calls and 3D virtual object exchange requires assistance of the most capable gateway. Collective application of deep meta data optimization requires cooperative management of the data flow overriding the application priorities selected for each dialog by partner's network rating policy. Multiparticipants' reward and punishment ratings distributed through the noise reduction system via an auxiliary embedded data rate manager, which can disallow emergence of non-cooperative agents, allows non-local distributed cooperation for optimization of any user defined data type flow through the network. Cooperation of intelligent agents, providing services for the same user, may also optimize their ongoing data flow, which can be redirected to prevent non-rentable activity.

9.5.1. Overview of Protocols

The traditional approach to communications and transport systems design consists of the definition of a set of interconnection and communication protocols. Defining interconnection and communications logic is tantamount to designing the transport architecture used by the application. On the contrary, this approach leaves to the application domain the full responsibility for implementing the logic of interconnections and communications. This lack of interconnection and communications automation leads to a number of problems when considering their application to the self-configuration area. Most of the problems stem from the fact that existing communications protocols are too mechanics and, therefore, unable to detect or automatically configure changes in the transport environment and user demands.

Protocol adaptation is a prerequisite for the automation of communication systems. The communication protocol concept is the key element that must be modified by self-configuring and self-managing algorithms. Thus, protocol adaptation will be the foundation for implementing the concepts defined in the previous chapters in a real heterogeneous and autonomous communications architecture. The definition of an adaptive protocol stack is a key point in the implementation of many of the user and operational requirements regarding self-managing or smart networks conceived in the last decade.

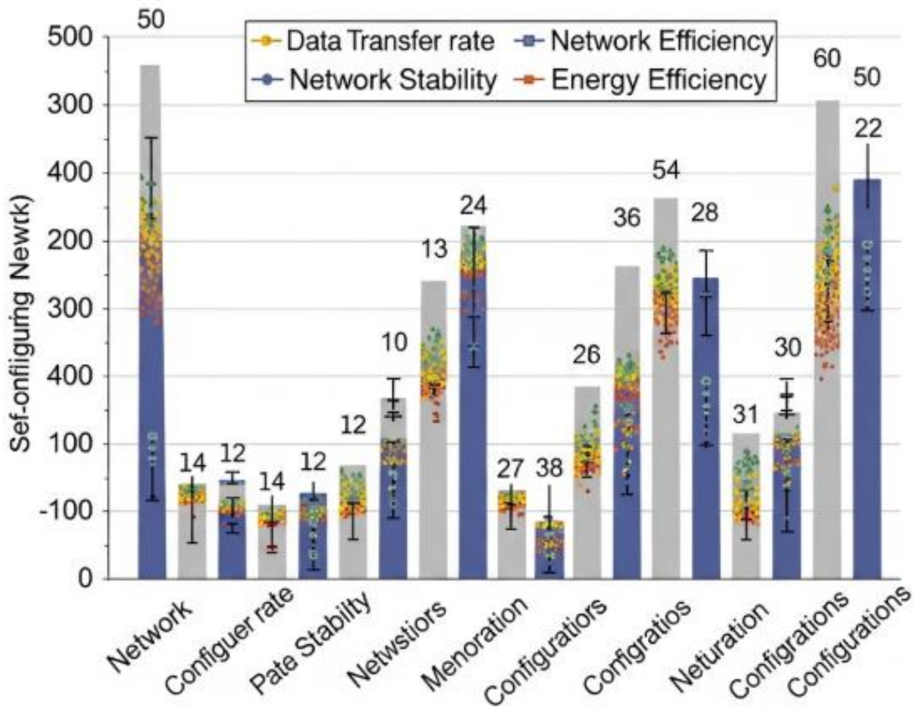

Fig 9.3: Embedded Intelligence in Autonomous Communication Systems

9.5.2. Protocol Adaptation in Self-Configuring Networks

Specific user applications impose functional and performance requirements on the services offered by communication networks. This in turn dictates specific design choices at both the system and protocol levels. Adaptivity and modularity are key features of a self-configuring network architecture. Given a set of user application requirements, a self-configuring communication system attempts to meet those requirements by dynamically (re)configuring the network and changing protocol behaviors as necessary. Network adaptivity extends to protocol adaptation, with the establishment of an efficient association between network services and protocol behaviors relevant for those services. In systems with a large number of functionalities and support for a variety of user applications, the principle of protocol adaptation becomes even more significant.

Current work on protocol adaptation is primarily at the level of design choices, such as choosing specific transport protocols and specific approaches for multicast communication. As an illustration, the switching mechanism is currently outside the control of an embedded communication system, as it is determined externally via explicit user commands or preset by the user/environment. In a self-configuring environment, a middleware calling subsystem would be called on demand to establish user-level objects, such as a session, and provide the requisite level of functionality. Such systems tend to be increasingly application-aware, and are well suited to a self-configuring environment.

9.6. Conclusion

The development of self-configuring networks and autonomous internetworking has begun with the automation of configuration of local and simple networks. In the task-oriented environment of a limited community with a finite number of users, autonomous communication systems based on embedded intelligence principles are already implemented. Adequate services have been defined and there are clear requests for a specialized and somewhat restricted set of services at the level of the community, to allow an efficient stand-alone operation of the self-configured internetworks. These limited networks resemble mobile local area networks. In the task-oriented community, users are identified as subjects of a limited database. Users have some identifiers associated with them. The communication is based on identifiers. Multicasting or broadcasting versus point-to-point communication is determined by the operation of the subjects or users in the database. Users are thus automatically connected to a self-organizing center of the limited database, communicating with the involved other users. For service provisioning, the fixed internetworks and points of presence act like service providers. Signaling is done using the address as a signal and control means.

It is important to note that the triggering control is the information activity of the users. The structure for both the software element and the hardware element is foreseen. The service structure is already defined. The elements of embedded intelligence are a communication endpoint, a specific software depending on the service being used, which configures a part of the system based on a centralized memory and a wireless cellular connection and portable terminals, which are connected for the limited time dedicated to the specific task execution and request control and memory access.

9.6.1. Future Trends

Future trends should be strongly influenced by the massive growth we are about to witness in the context of the amount of resources and devices connected to the network at different levels. The larger the amount of devices connected, the more difficult, but also stimulating, is the task regarding network management and operation, symbolizing a strongly desired lower cost of ownership. New media and access technologies are about to appear, each of them suitable for a very limited class of service requirements. It is very likely that new services requiring very low costs, very low latency, high mobility and rapidly changing topological configurations will appear.

In the near future, specialists in the areas of wireline and wireless computer networks should be aware of the challenges posed by such a rich environment. It will be imperative to modify the way we design and develop network devices, nodes, access networks, internetworks and systems, as well as the services, protocols and algorithms that control their behavior. All devices should be enabled to acknowledge their surrounding conditions and adjust corresponding resource parameters in order to reach a network-wide optimal operation. Autoconfiguration and self-organization at the network edge, however, requires a careful design of the underlying self-configuration network protocols and mechanisms employed. Bottom-up design approaches. Behavioral approaches to self-configuration, learning agents, are key elements necessary for both performance and stability. Low level device autoconfiguration, learning and mobility mechanisms should be built and work in a layered, cohesive manner to construct a global self-configuration capability. Self-configuration capabilities incorporated in the network primitives should guarantee the stability and suitability of the global system behavior. An enabling technology for both new service requirements and the self-configuration capability is based on a new look at the network control and management application area. Behavioral approaches to autonomous systems design, intelligent agents, and adaptive network management and control are key elements, necessary for achieving performance and stability.

References

Han, S., Mao, H., & Dally, W. J. (2016). Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. ICLR.

Li, H., Dechter, A., & Roy, K. (2020). Hardware-aware neural architecture search: Modeling and optimization. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 39(11), 3223–3236.

Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proceedings of the IEEE, 105(12), 2295–2329.

Zhang, X., Wang, Y., Zhang, Y., Lin, Y., & Shi, Z. (2019). DNNBuilder: An Automated Tool for Building High-Performance DNN Hardware Accelerators for FPGAs. IEEE/ACM Transactions on Computer-Aided Design of Integrated Circuits and Systems, 38(10), 1959–1972.

Chen, Y., Emer, J., & Sze, V. (2016). Eyeriss: A Spatial Architecture for Energy-Efficient Dataflow for Convolutional Neural Networks. ACM SIGARCH Computer Architecture News, 44(3), 367–379.