Applying and validating AI-based QoS profiling for NGN user terminals

F. Weber[1], W. Fuhrmann[2], U. Trick[1], U. Bleimann[2], and B. Ghita[3]

[1]Research Group for Telecommunication Networks, University of Applied Sciences Frankfurt/M., Frankfurt/M., Germany
[2]University of Applied Sciences Darmstadt, Darmstadt, Germany
[3]Centre for Security, Communications and Network Research, University of Plymouth, Plymouth, UK
e-mail: weber@e-technik.org

Abstract

This paper addresses the identification of NGN (Next Generation Networks) user terminals experiencing similar QoS (Quality of Service) conditions. This approach aims on the reduction of network traffic resulting from comprehensive QoS monitoring. An ART 2 ANN (Adaptive Resonance Theory 2 Artifical Neural Network) has been evaluated for the comparison and classification of sequences of consecutive jitter (delay variation) values experienced by packets of simultaneous multimedia over IP data streams. Further on, a procedure is introduced supporting the classification process through automated result validation.

Keywords

Next Generation Networks, Quality of Service, Artificial Neural Networks, ART 2

1. Introduction

Quality of Service (QoS) is one of the key features of the standardised NGN concept (ITU-T Y.2001, 2004), (ETSI TR 180 000, 2006). Unfortunately, as stated in (Park and Kang, 2005) and (Weber *et al.*, 2007), the active control of network resources within the IP transport network results in a considerable amount of resource management traffic. In order to address this issue, an integrated framework for comprehensive QoS control in SIP-based NGN has been introduced in (Weber *et al.*, 2008, 1). One key characteristic of this framework is the continuous collection of information on the QoS experienced by any NGN user terminal.

As demonstrated in previous research work (Weber *et al.*, 2009), NGN user terminals experiencing similar QoS conditions can be virtually grouped. By monitoring the QoS conditions in one user terminal of each group, conclusions can be drawn on the QoS experienced by any user terminal being a member of the respective group. This results in a comprehensive but resource-saving QoS monitoring concept. Further on it was found that ANN can be used to assign NGN user terminals to virtual groups (referred to as QoS profiling).

This paper proposes a method for QoS profiling and a self-contained validation procedure, based on ANNs of the type ART 2 (Adaptive Resonance Theory 2) introduced in (Carpenter and Grossberg, 1987). Initial tests have been accomplished and the result trends are introduced within this paper.

2. Next Generation Networks (NGN) and Quality of Service (QoS)

In 2004, the ITU-T (International Telecommunication Union – Telecommunication Standardization Sector) released its definition of NGN in (ITU-T Y.2001, 2004). According to (ITU-T Y.2001, 2004), (ETSI TR 180 000, 2006), and (Trick and Weber, 2009) the term NGN stands for a telecommunication network concept that can be characterised by a number of key features including, amongst others, "Packet-based data transport" and "Quality of Service support". Although the term "Packet-based data transport" does not refer to any particular technology or protocol, IP (Internet Protocol) is the most likely network protocol choice for an NGN environment according to (Trick and Weber, 2009). The use of SIP (Session Initiation Protocol) for NGN service provisioning and signalling is widely accepted, and also suggested in (ETSI ES 282 001, 2008).

2.1. QoS for real-time telecommunication services

For services provided within telecommunication networks, the term QoS has been defined as the "collective effect of service performance which determine the degree of satisfaction of a user of the service" (ITU-T E.800, 1994). According to (ITU-T Y.1291, 2004), for packet-based media data transport (which is given in NGN), the quality of a real-time based telecommunication service as experienced by a service user directly depends on the network performance of the respective transport network. In (Gozdecki *et al.*, 2003) the network performance of an IP transport network is characterised by the packet loss ratio, the transfer delay, and the transfer delay variation (jitter). These network performance parameters substantially influence the QoS of a real-time communication service as experienced by its users.

2.2. Integrated framework for comprehensive QoS control in SIP-based NGN

The authors proposed in a previous study a framework for QoS control, aiming to address the scalability issues related to QoS provision in SIP-based NGN, as described in (Park and Kang, 2005) and (Weber *et al.*, 2007). Within this framework approach (Weber *et al.*, 2008, 1), all action required for the control of the QoS affecting media sessions is performed within the NGN service stratum (i.e., cross-strata communication is avoided). Therefore, the framework has to be provided with an integrated mechanism for the collection of information on the QoS affecting any ongoing and future media session.

This information consists of delay, jitter, and packet loss values affecting the packets of a respective media data stream. The information is best collected close to the receiving user terminal of the respective data stream to consider the sum of effects appearing on the entire network path between sender and receiver. In order to minimise the resulting QoS monitoring traffic, only selected user terminals (representing a QoS reference point) are queried for information on the QoS conditions experienced by ongoing media sessions. This requires the identification of virtual groups of user terminals whose members experience similar QoS conditions, and hence, can be represented by a specific reference point. The assignment of user terminals to their respective virtual group, resulting from the comparison of QoS conditions experienced, is proposed to be performed by the use of an Artificial Neural Network (ANN) to allow for an improved real-time processing behaviour. The principle of assigning media streams to virtual groups (or classes, respectively) by the use of an ANN is introduced in section 4 of this paper. A detailed description of the overall framework functionality of QoS information collection is provided in (Weber *et al.*, 2008, 1) and (Weber *et al.*, 2008, 2).

3. ART 2 Artificial Neural Networks

Artificial Neural Networks (ANNs) are used in numerous technical applications in order to perform complex tasks such as pattern recognition (or pattern classification), function approximation, prediction/forecasting, optimisation, content-addressable memorising, cybernetics, as well as clustering/categorisation. The latter application is denoted as unsupervised pattern classification in (Jain et al., 1996).

ART 2 (Adaptive Resonance Theory) neural networks can be described as unsupervised-learning neural networks with the ability to compare analogue continuous value sequences with the objective to classify the sequences by their similarities (Carpenter and Grossberg, 1987). This is performed by self-organisation of stable recognition codes generated from the input value sequences. An input sequence, also referred to as a pattern, is interpreted as an n-dimensional vector by an ART network, where n is the number of values comprised by the respective input pattern.

An ART 2 ANN provides n input units and m output units, the latter of which represent m individual output classes. If an arbitrary number of n-dimensional patterns is presented to an ART 2 network, after a predefined number of learning cycles, the network tries to map each pattern to one of m output classes by accomplishing a multi-step comparison process for each pattern. Patterns showing typical similarities are to be assigned to the same output class. For the comparison and classification of the patterns, the ART 2 ANN interprets every pattern as an n-dimensional vector, where n refers to the number of input values per pattern.

A number of setup parameters are provided by ART 2 ANN, of which the vigilance parameter ρ is the most effective. ρ represents a selectable threshold for the deviation $||r||$ of two n-dimensional vectors u and cp (see equation (1)), where u represents the candidate pattern to be currently classified. Vector cp represents the ART2-internal pattern image of a certain class. With e=0, it is obvious from equation (2) that a reset event is triggered when $||r|| < \rho$. The reset event causes the ART2-internal resumption of the classification process of the respective pattern represented by u_i, excluding the respective class represented by cp_i.

$$r_i = \frac{u_i + cp_i}{e + ||\mathbf{u}|| + ||c\mathbf{p}||} \qquad (1) \qquad\qquad \text{Reset} \Leftrightarrow \frac{\rho}{e + ||\mathbf{r}||} > 1 \qquad (2)$$

Further details on the theory of ART 2 neural networks can be found in (Carpenter and Grossberg, 1987).

4. AI-based QoS profiling for NGN user terminals

As mentioned in section 1, the term ‘QoS profiling’ refers to the virtual grouping of NGN user terminals by QoS conditions encountered. The reason for applying QoS profiling is the reduction of network traffic resulting from comprehensive QoS monitoring.

The integrated NGN QoS control framework briefly introduced in section 2.2 provides a centralised unit for the rating of QoS conditions. It is assumed that all NGN user terminals associated with a virtual group encounter similar QoS conditions. Hence, it is sufficient to choose one group member as the group’s reference point and, subsequently, gather and rate QoS condition information from this reference point only. This information is assumed to represent the QoS experienced by any member of the respective virtual group.

Note that at least one reference point has to be chosen per existing virtual group by a selection process, further described in (Weber *et al.*, 2008, 2).

Each group’s selected reference point is queried to continuously provide QoS information to the framework’s centralised QoS rating unit. The QoS information gathered is used in two ways:

- to rate the QoS experienced by user terminals (and, hence, users) associated with the respective virtual group (as described in (Weber *et al.*, 2008, 1), our framework can react upon deteriorating QoS conditions with the objective of improvement).
- as reference information representing the respective virtual group (required for the process of assignment of further new user terminals to one of the existing virtual groups).

The latter process is described within the following sections.

4.1. Comparing QoS characteristics

In order to group NGN user terminals by their QoS characteristics encountered, comparable data have to be available, representing the QoS characteristics of each existing virtual group and of the user terminal to be assigned to one of the groups. According to our experiments introduced in (Weber *et al.*, 2009), the QoS characteristics of packet streams were found to be best represented by jitter values (see section 2.1) because of their susceptibility to changes within the IP network load utilisation. Note that the jitter data to be compared have to be synchronised in time, as the jitter characteristic experienced by an user terminal is represented by an ever-changing value progression. Figure 1 shows exemplary jitter characteristics, synchronously monitored in five different NGN user terminals.

In this example, four jitter characteristics were provided by NGN user terminals serving as reference points (grey-coloured characteristics Ref. 1 … Ref. 4), while the fifth characteristic was monitored in a user terminal which has to be assigned to one of the groups represented by the reference characteristics. Comparing the jitter characteristics given in Figure 1, considering the full time interval shown, it can be determined that the user terminal to be assigned (bottom line, black) is most likely a member of the group represented by Ref. 3 (third grey line from the top).

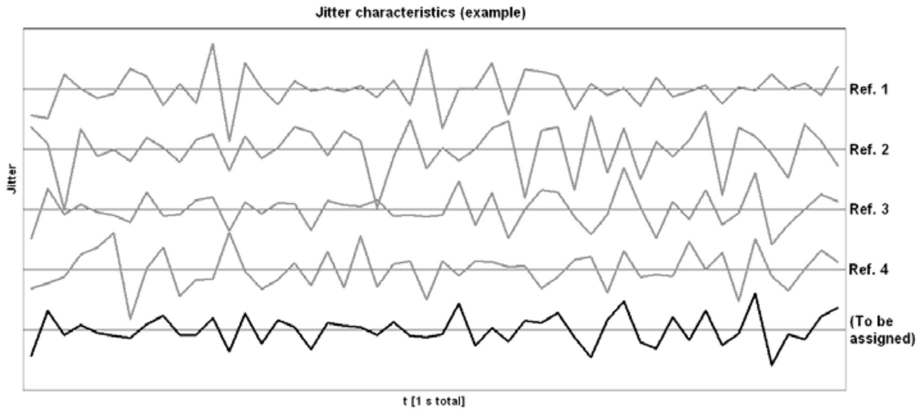

Figure 1: Jitter characteristics, monitored in NGN user terminals

4.2. Applying an ART 2 ANN for the grouping of NGN user terminals

For the automated comparison and grouping of jitter characteristics, an ART 2 ANN (see section 3) has been selected, because of its suitability in comparing and classifying analogue value sequences. The ART 2 ANN is provided with n input cells and m output cells. In our application, n refers to the number of discrete values to be considered within the ART 2 classification run. Hence, if a number of 50 consecutive jitter values is chosen to be considered (which corresponds to a media sequence of one second for typical VoIP (Voice over IP) calls established with, e.g., G.711 codec), the ART 2 ANN has to be provided with 50 input cells.

The number of output cells m depends on the number of existing virtual groups to be considered within the ART 2 classification run. Thus, if four reference jitter characteristics are available, each representing the QoS conditions of one existing virtual group, the ART 2 ANN has to be provided with four output cells.

As described in section 3, a sequence of data to be classified within an ART2 classification run is referred to as a pattern. Hence, a set of value sequences, each representing one group, is referred to as a pattern set. Within the following, a jitter value sequence is named a pattern.

All jitter patterns to be considered (one reference pattern per existing virtual group plus the pattern to be classified, referred to as 'Pattern Under Test' (PUT)) are provided to the ART 2 ANN as a pattern set. Within one classification run, the ANN performs a number of internal comparison and weighting tasks, involving an unsupervised learning process, resulting in a stable state of classification. When this state is reached, each pattern has been either assigned to exactly one class, or it could not be assigned to any class. Note that several patterns within one pattern set could also be assigned to the same class if they showed effectual similarities.

Also note that the result of an ART 2 classification run is strongly influenced by the choice of the ART 2 vigilance parameter ρ (see section 3), providing a value range of ρ = [0.7…1.0]. If ρ is chosen relatively low, the ANN shows a more tolerant classification behaviour. Hence, the classification is potentially imprecise. If ρ is chosen relatively high, the ANN shows a strict classification behaviour, coming along with the risk of being unable to identify similarities among patterns within the respective pattern set (see section 3 for details on the function of ρ). Thus, the result of an ART 2 classification can only be considered reliable if the value chosen for ρ is suitable for the respective classification problem.

Note that the most suitable value for ρ, amongst others, depends on the texture of the patterns to be classified, and hence, can generally not be pre-defined. In (Rayón Villela and Sossa Azuela, 2000) a method is introduced for the determination of ρ in ART 2 ANN. However, as this approach is optimised on the best possible discrimination of patterns only, this approach is not suitable for the process of determining reference patterns and, at the same time, assigning another pattern to one of the references.

4.3. A validation procedure for the ART 2 classification

Because for our purpose it is understood that the reference patterns must be distinguishable from each other (as they are known to represent different QoS characteristics), we can use this fact for approximating the most suitable value of ρ. Therefore a number of classification runs are performed with the same pattern set, but with varied ρ (note that within the following, performing multiple classification runs with the same pattern set is referred to as one classification process). Table 1 shows an example for the influence of ρ on the result of ART 2 classification runs, and for how the most suitable value ρ can be evaluated.

Classification Process No.	**Vigilance (ρ)**	**Pattern 1** (Ref. 1)	**Pattern 2** (Ref. 2)	**Pattern 3** (Ref. 3)	**Pattern 4** (Ref. 4)	**Pattern 5** (PUT)
1	0.85	1	1	1	2	1
2	**0.934375**	**1**	**2**	**3**	**4**	**3**
3	0.94375	1	2	3	4	0

Table 1: Example for the influence of ρ on ART 2 classification results

For the classification runs considered, in run

- no. 1 (ρ = 0.85), all patterns except pattern 4 were assigned to the same class (class 1). Pattern 4 was assigned to a separate class (class 2). Hence, as the classification result is too imprecise to distinguish among the reference patterns (patterns 1…4), the value for ρ has been chosen too low.
- no. 2 (ρ = 0.934375), all patterns representing references (patterns 1…4) were assigned to separate classes (classes 1…4), while the PUT (pattern 5) was assigned to class 3. As pattern 5 was assigned to the same class as pattern 3, it is evident that the user terminal represented by pattern 5 belongs to the same group as the user terminal serving as reference 3.

However, ρ might still be chosen too low, as the classification behaviour of the ART 2 would perhaps be more exact with a higher value of ρ.

- no. 3 (ρ = 0.94375), all patterns representing references (patterns 1…4) were assigned to separate classes (classes 1…4), while the PUT (pattern 5) could not be assigned to any of these classes (showing “0”). Hence, either the user terminal represented by pattern 5 shows a different jitter characteristic than the user terminals serving as references 1…4 (in this case, a new virtual group is established), or ρ was chosen too high to identify similarities between pattern 5 and another pattern.

Thus, for the exemplary classification process shown in Table 1, the best suitable value for ρ must be in the range of (0.934375…0.94375). In order to further pinpoint the best suitable value for ρ, further classifications have to be run with ρ values taken out of the given range. However, if these further classification runs do not disprove the provisional result evident from Table 1, the PUT (pattern 5) can be said to be classified into the group represented by reference 3. This is due to the result of classification run no. 2, in which the reference patterns are distinguished from each other and, at the same time, the PUT could be assigned to one of them.

In order to define a consistent procedure for the validation of ART 2 classification run results, the following cases of possible results are distinguished. Each case provides a statement regarding the valuation of the applied ρ value, aiming towards a most suitable ρ. This statement provides the basis for the choice of ρ to be set for the subsequent classification run. Following these statements, a most reliable assignment of the PUT can be provided, and a most suitable value for ρ can be determined.

- Case a) If at least one reference pattern could not be assigned to any class number, the classification process was performed too strict. Hence, ρ was chosen *too high*.
- Case b) If two or more reference patterns share the same class number, the classification process was performed too loose. Hence, ρ was chosen *too low*.
- Case c) If all reference groups could be distinguished correctly from each other and if a class number was assigned to the PUT, this class number is a potential group candidate. Hence, the classification process might have been performed well, but it also might have been performed too loose. Hence, ρ was potentially chosen *too low*.
- Case d) If all reference groups could be distinguished correctly from each other and if no class number was assigned to the PUT, this could have been caused by two different situations. In any case, ρ was potentially chosen *too high*.
    - The classification process might have been performed well. In this case, the PUT simply does not match any of the reference groups and, hence, is the first recognised representative of a new group. We assumed that this was true if this PUT could not be assigned to a reference pattern within another classification run performed.

- The classification process might have been performed too strict to identify similarities between the PUT and any of the reference patterns.

In order to determine a most suitable value for ρ, multiple classification runs have to be performed based on the same pattern set. In each run, ρ has to be adapted subject to the outcome of the previous run (cases a) … d)). Regarding the adaption, we suggest to follow the rules below, resulting in a statistically optimised number of required classification reruns (i = number of classification runs performed with the same pattern set. See case explanations above for the meaning of "too low/high").

$\rho_{(0)min} = 0.7$; $\rho_{(0)max} = 1.0$ $\qquad$ $\rho_i = (\rho_{(i)max} + \rho_{(i)min}) / 2$

If $\rho_{(i-1)}$ = "*too low*": $\rho_{(i)min} = \rho_{(i-1)}$; $\rho_{(i)max} = \rho_{(i-1)max}$

If $\rho_{(i-1)}$ = "*too high*": $\rho_{(i)max} = \rho_{(i-1)}$; $\rho_{(i)min} = \rho_{(i-1)min}$

This classification process is continued in an automated manner until a most suitable value for ρ is determined (i.e., until the assignment of the PUT to either one specific class or to no class has been verified within a preselected number of classification runs, or until a most suitable value for ρ could be narrowed down to a predefined range of accuracy).

The results of an ART 2-based jitter pattern classification process can be further verified by repeating the process for jitter patterns originating from the same user terminals, but which represent a later time interval of the same respective communication situations. After having performed several classification processes, the provisional results can be compared and the most likely final result can be determined.

5. Tests

The classification validation procedure described in section 4.3 has been exemplarily tested in a proof-of-concept manner. Initial result trends are introduced within this section.

5.1. Test infrastructure

Based on the ns-2 network simulator, a SIP-based NGN architecture has been set up, allowing for the simulation of different communication scenarios. The architecture consisted of one core network and four access networks, to each of which different numbers of user terminals were connected. Upon session initiation, media flow packets (simulating VoIP calls with G.711 codec) were bidirectional exchanged in a peer-to-peer manner between the user terminals. By varying the distribution of parallel calls among the user terminals connected to different access networks, different numbers of virtual user groups were set up, each providing varying QoS conditions. Further information on the theory of generating virtual groups can be found in (Weber *et al.*, 2008, 2).

By the use of the ns-2 trace function, all media packets were recorded and time-stamped at their respective receiving user terminals. The collected data were post-processed to extract the per-packet inter-arrival jitter of each media flow. For each virtual user group, one user terminal was randomly chosen whose jitter characteristic served as the respective group's reference characteristic. Finally, the synchronised jitter sequences were bundled as pattern sets and provided to the ART2 ANN (see section 4.1 and 4.2), and the classification procedure introduced in section 4.3 was applied.

5.2. Result trends

Table 2 shows the result trends of the initial proof-of-concept tests performed for the evaluation of AI-based QoS profiling.

Scenario	No. of virtual groups included	Accuracy of classification
a)	4	93%
b)	8	71%
c)	10	69%

Table 2: Result trends of initial tests of the jitter classification

As shown in Table 2, within the initial tests, three different communication scenarios (a) … c)) were considered, differing in the number of virtual groups included. It is observed that the classification accuracy decreases from scenario to scenario with the increase of the number of virtual groups. Although it is evident that the principle of AI-based QoS profiling, applied in conjunction with the evaluation procedure introduced within this paper, results in a suitable degree of accuracy. However, it is obvious that this procedure has to be further improved to allow for more virtual groups to be considered while applying the procedure introduced within this paper.

One improving approach could be to detect the similarity between a respective PUT and each group reference separately, applying an ART 2 ANN and the validation procedure as described in sections 4.2 and 4.3. In this case, the degree of similarity between the PUT and each respective reference pattern can be quantified as a reciprocal function of ρ. In case of ambiguity, a pattern set can be arranged comprising the PUT and a number of the most similar reference patterns, and the procedures described in sections 4.2 and 4.3 can be applied. This will result in a more accurate classification.

6. Conclusion

Within this paper, the application of Artificial Neural Networks of the type ART 2 for QoS profiling in Next Generation Networks has been described. An automated validation procedure has been introduced, assisting the classification process through evaluation of intermediate results. This procedure has been tested in a proof-of-concept manner in a network simulation environment and the result trends are provided within this paper. It was found that the accuracy of the introduced classification and validation procedures depends on the number of QoS profiles that can be differentiated within a considered communication situation. A possible approach has been named to avoid the consequences of this issue.

7. References

Carpenter, G.A. and Grossberg, S. (1987). ART 2: "Self-organization of Stable Category Recognition Codes for Analog Input Patterns", Applied Optics, 26 , pp. 4919-4930

ETSI ES 282 001 V2.0.0 (2008), ETSI Standard, "NGN Functional Architecture", ETSI TISPAN

ETSI TR 180 000 V1.1.1 (2006), Technical Report, "NGN Terminology", ETSI TISPAN

Gozdecki, Janusz; Jajszczyk, Andrzej and Stankiewicz, Rafal (2003), "Quality of service terminology in IP networks", Communications Magazine, Volume 41, Issue 3, pp. 153-159, IEEE

ITU-T E.800 (1994), Recommendation, "Terms and definitions related to Quality of Service and Network Performance including dependability", ITU-T

ITU-T Y.1291 (2004), Recommendation, "An architectural framework for support of Quality of Service in packet networks", ITU-T

ITU-T Y.2001 (2004), Recommendation, "General overview of NGN", ITU-T

Jain, A.K.; Jianchang Mao; Mohiuddin, K.M., (1996), "Artificial neural networks: a tutorial", Computer, vol.29, no.3, pp. 31-44, Mar 1996

Park, Juyoung and Kang, Shin Gak (2005), "QoS Architecture for NGN", Advanced Communication Technology, ICACT 2005, pp. 1064-1067, IEEE

Rayón Villela, P.; Sossa Azuela, J. H. (2000), "A Procedure to Select the Vigilance Threshold for the ART2 for Supervised and Unsupervised Training", LNCS, vol. 1793/2000, pp. 389-400, 2000, Springer

Trick, Ulrich and Weber, Frank (2009), SIP, TCP/IP und Telekommunikationsnetze (4th edition), Oldenbourg, Munich, Germany, ISBN: 978-3-486-59000-5

Weber, F; Fuhrmann, W.; Trick, U.; Bleimann, U. and Ghita, B. (2007), "QoS in SIP-based NGN – state of the art and new requirements", Proceedings of the third collaborative research symposium on Security, E-learning, Internet and Networking SEIN 2007, pp. 201-214, Information Security & Network Research Group – University of Plymouth, Plymouth, ISBN: 978-1-8410-2173-7

Weber, F.; Fuhrmann, W.; Trick, U.; Bleimann, U.; and Ghita, B.V. (2008), 1, A Framework for Improved QoS Evaluation and Control in SIP-Based NGN, Proceedings of the Seventh International Network Conference (INC2008), Plymouth, UK, 8-10 July, pp. 27-37, 2008

Weber, F.; Fuhrmann, W.; Trick, U.; Bleimann, U.; and Ghita, B.V. (2008), 2, "Selection of QoS Monitoring Points in a New QoS Control Framework for SIP-Based NGN", Proceedings of the Fourth Collaborative Research Symposium on Security, E-learning, Internet and Networking (SEIN 2008), Wrexham, UK, ISBN: 978-1-84102-196-6, pp. 176-185, 2008

Weber, F.; Fuhrmann, W.; Trick, U.; Bleimann, U.; and Ghita, B.V. (2009), AI-based QoS profiling for NGN user terminals, Proceedings of the Third International Conference on Internet Technologies & Applications (ITA09), Wrexham, UK, 8-11 September 2009