Characterization of the Link Layer Service Capacity of Adaptive Air Interfaces with Imperfections

Ulrico Celentano, and Savo Glisic, Senior Member, IEEE

Abstract—A model for the service capacity at link layer, for a wireless, link-adaptive system is presented. The model includes imperfections in the adaptation chain (estimation error, estimation delay, acquisition error), and implementation implications (switching hysteresis). Transceiver characteristics and imperfections are independently represented by separate matrices. The dependence of effective capacity $R_c$ on the impairments is discussed using analytical, numerical, and simulations results. Examples show that the above effects should not be neglected in realistic performance analysis at upper layers.

Index Terms—Adaptive systems, Cross-layer, Effective capacity, Goodput, Impairments, Information rates, Link layer, Reconfigurable communication networks.

I. INTRODUCTION

SEVERAL phenomena affect a wireless link capacity. Primary effects are essentially propagation properties of the environment and mobility of terminals. Secondary effects include dynamic adaptation of transceiver techniques to actual conditions. Transceivers reconfiguration, introduced to compensate for the primary impairments, must be operated at both sides, consistently. Imperfections, intrinsic in real adaptive systems, affect system performance. Those aspects must be characterized with sufficient accuracy but still with a simple and tractable model that can be used in the analysis of the higher network layers.

For design and performance analysis of upper layer protocols, it is important to characterize the actual link service capacity. In this paper, we look at the portion of the protocol stack below medium access control (MAC) layer [1]. The same portion of the protocol stack is covered in a link layer model, called effective capacity link model [2], which models directly few link layer parameters used in queuing analysis, without including imperfections of the physical layer or adaptation in link layer. In [3], a model for packet losses is included, taking into account physical channel, modulation and channel coding, and some other functions of the data link layer, but only for transceivers with fixed structure.

In this paper, a model for the service capacity of wireless adaptive links is presented. The effective quantity of service provided at upper layers is measured by the goodput. The average goodput is expressed in compact form as $\bar{R}_c = r^T \text{diag}(Y \Theta^T)$, see Sect. IV. The dependence of $R_c$ on the impairments is discussed in Sect. V using analytical, numerical, and simulations results. Illustrative examples show that the imperfections, modeled in Sect. III, should not be neglected for realistic performance analysis at upper layers.

II. EFFECTIVE CAPACITY OF THE IDEAL ADAPTIVE LINK

In link adaptive systems, as a response to the received signal quality $\gamma$, the transceiver configuration $K_{\text{trx}}$ (PHY-mode or mode) is changed so that a performance metric $e$ is kept between certain boundaries: $e(K_{\text{trx}}, \gamma) \in \{ \text{acceptable values} \}$. The received signal quality can be measured, e.g., with the signal to interference-plus-noise ratio (SINR). In the following, we denote with $\gamma$ a generic metric for the quality level. An example of metric $e$ is the error rate, but other metrics may be used instead.

Link adaptation strategies may include adaptive modulation and coding (AMC), power control, or both [4]. As we will see, net throughput $r$ and residual error rate $e$ summarize the contributions of the configuration to the effective link capacity. In our framework, a mode may be associated with a wider set of transceiver configuration parameters, as enabled by software-defined radio, as long as that mode is represented by an $(r, e)$ pair. Non-adaptive systems, having a single, unchangeable transceiver configuration (modulation scheme, channel coding scheme, etc.), are called here fixed systems.

The cardinality of the modes set $\mathcal{M} = \{ M_0, M_1, \ldots, M_{M-1} \}$. $M = |\mathcal{M}|$, is typically small. For example, $M$ is up to six in IEEE 802.16a, seven in HSDPA of UTRA (High Speed Downlink Packet Access of Universal Mobile Telecommunication System Terrestrial Radio Access) and ETSI (European Telecommunications Standards Institute) HiperLAN2, and eight in IEEE 802.11a.

The domain of the signal quality metric is divided into $M$ contiguous, non-overlapping regions: $S = \{ S_0, S_1, \ldots, S_{M-1} \}, S_i \leftrightarrow [\gamma_i, \gamma_{i+1})$ [1]. Each region is associated with a mode.

The goodput is defined as the net bit rate, thus excluding overhead of headers, redundancy of forward error correction coding, and errors that occur in the communication channel. Adapting an expression for the link service rate from [5], we write the goodput for an ideal fixed system as $R_c(t) = [1 - e(t)] r(t)$, where $r(t)$ is the informative bit transmit rate or net throughput, and $e(t)$ is the residual error rate at reception. For an ideal reconfigurable system, each state is associated with, and characterized by, its transmit rate and error rate: $S_i \Rightarrow M_i \Rightarrow r_i, e_i$. 

Manuscript received July 4, 2004; revised September 30, 2005; accepted February 24, 2006. The editor coordinating the review of this paper and approving it for publication is K. K. Leung. This research was supported by the Finnish Funding Agency for Technology and Innovation (Tekes), Nokia, and the Finnish Defence Forces, Elektrobit and Instrumentointi. Part of this paper was presented at the 16th Annual IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Berlin, Germany, Sept. 11–14, 2005.

U. Celentano is with the Centre for Wireless Communications, University of Oulu, Oulu, Finland (e-mail: ulrico.celentano@ee.oulu.fi).

S. Glisic is with the Telecommunication Laboratory, University of Oulu, Oulu, Finland (e-mail: savo.glisic@ee.oulu.fi).
The informative bit transmit rate depends on the transceiver configuration (mode). For mode $M_j$, we have $r_j = \eta_h \eta_e k_j / T_s$, where $\eta_h = L_i / L_d$ is the header efficiency, $L_d = L_i + L_h$ is the total size of payload plus header, $\eta_e = L_d / (L_d + L_o)$ is the channel coding code rate, including redundancy $L_o$, $k_j$ is the number of bits per symbol of the $j$-th mode modulation scheme, and $T_s$ is the symbol duration. A received packet is valid if no error is present or all errors have been corrected. Therefore, the residual error rate after reassembling $e$ has the same value of the packet error rate (PER): $e = p_E$.

The effective capacity of the link is obtained by averaging the goodput over the modes:

$$\bar{R}_c|_{\text{ideal}} = \sum_{i=0}^{M-1} [1 - e_i] r_i \pi_i,$$

where $\pi_i$ is the probability of the channel being in the $i$-th state and using the $i$-th mode.

An adaptive system has a reference quality and a control. The control (actual measured quality) can be defined at the receiver (closed loop) or at the transmitter (open loop). In either case, it is impossible for the transmitter to know what the actual quality will be at the reception instant and site.

### III. Model of Imperfections

The properties of the transceivers configuration set-up errors depend on the particular system. In this paper, it is out of scope the study of specific algorithms, but rather we build a flexible model open to generalizations.

The metric used for mode switching can be expressed as $\tilde{\gamma}(t) = \gamma(t - \tau_e) + \varepsilon_e + \varepsilon_f$, where $\gamma$ and $\tilde{\gamma}$ are the true and the estimated value, respectively, $\varepsilon_e$ is the estimate error, $\varepsilon_f$ is the acquisition error, and $\tau_e$ is the estimation delay. At estimation time $t$, the true channel quality metric falls in the $k$-th region, $S(t) = S_k$, and the metric is estimated: $\tilde{S}(t) = S_h$. The estimator or mode selector may be implemented at transmitter or receiver side. The estimation or command is sent to the other side, and $M_i$ is finally acquired, and used for transmission (with closed loop) or reception (with open loop). During the estimation delay $\tau_e$, the channel quality metric may move to its effective region, $S(t + \tau_e) = S_j$. The final effect is that mode $M_i$ is used at transmission time, when the channel is in effective state $S_j$. The probability of this event can be written as $P(M_i, S_k, M_j, S_j)$. We assume the estimation process and the channel process independent. (This relates mainly to the estimation algorithm.) We assume also that the acquisition channel is independent of the direct channel. (This relates to the adaptation scheme, duplexing scheme, MAC frame structure, terminals speed, etc.) Under these assumptions, we can write

$$P(M_h, S_k, M_i, S_j) = P(S_k) P(M_h | S_k) P(M_i | M_h) P(S_j | S_k).$$

The three last terms in (2) are treated separately in the sequel.

The mode estimation probability matrix $H(e) = \{h_{ck}\} \in \mathbb{R}^{M \times M}$ models the effects of the imperfections in the estimation process due to noise. The probability of selecting mode $M_h$ with the true channel in state $S_k$ at estimation time is $h_{ck} = P\{M_h | S_k\}$. This is the second term in (2).

The mode acquisition probability matrix $H(f) = \{h_{ik}\} \in \mathbb{R}^{M \times M}$ models the probability of errors in the exchange of set-up information between transmitter and receiver. The set-up information is the identifier of the mode or, equivalently, of the related signal quality region. We define: $h_{ik} = P\{M_i | M_h\}$. This represents the third term in (2).

The delayed channel transition matrix $H(d) = \{h_{kj}\} \in \mathbb{R}^{M \times M}$ models the effects of estimation delay. The probability that given the true channel was in state $S_j$ at estimation time, during transmission delay (after $\tau_e$) the channel is in state $S_j$ is $h_{kj} = P\{S(t + \tau_e) = S_j | S(t) = S_k\}$. This represents the last term in (2).

Although not exactly an imperfection, the following aspect is also related to implementation of link adaptation in a real system. To avoid possible too frequent mode switching around thresholds, switching hysteresis can be introduced. Two distinct values for falling, $\gamma^-$, and rising, $\gamma^+$, thresholds are defined: $\gamma^+ = \gamma^- + \varphi^+ \geq 0$. The width of the hysteresis region is $\varphi^+ + \varphi^-$. Margins $\varphi^-\gamma$ can be equal, or set so that the cumulative density function in the hysteresis regions is the same at both sides: $\Delta^{(\gamma^-)}_{CDF} = \int_{\gamma^-}^{\gamma^+} p_\gamma(\gamma)d\gamma = \int_{\gamma^-}^{\gamma^+} p_{\gamma^-}(\gamma)d\gamma = \Delta^{(\gamma^+)}_{CDF}$. Alternatively, an effective definition could be done imposing an upper limit on the probability of spurious switching in a given time.

### IV. Effective Capacity of Imperfect Adaptive Links

The service capacities can be represented in (1) by a vector because the effective state and chosen mode by assumption coincide: $S_i \leftrightarrow M_i$. In case of imperfect systems, the bit rate depends on the used mode whereas the error rate depends on both the used mode and the effective state. The following definitions are needed to extend the expression to imperfect systems.

The throughput vector $\mathbf{r} = \{r_i\} \in \mathbb{R}^M$ is defined as $\mathbf{r}^T = [r_0, \ldots, r_{M-1}]$ with $r_i$ defined above. The normalized goodput matrix is $\mathbf{Y} = (y_{ij} = (1 - e_{ij})) \in \mathbb{R}^{M \times M}$, where $e_{ij}$ denotes the error rate when the mode $M_i$ is used and the channel is in effective state $S_j$ (see Sect. III). Matrix $\mathbf{Y}$ expresses the useful bit rate normalized to the transmit rate in each case. The sojourn probabilities in the signal quality regions are denoted with the state probability vector $\mathbf{\pi} = \{\pi_j\} \in \mathbb{R}^M$.

The effective capacity, in presence of the imperfections modeled in Sect. III, is

$$\bar{R}_c = \sum_{i=0}^{M-1} h_{ti}^{(f)} r_i \sum_{j=0}^{M-1} (1 - e_{ij}) \sum_{k=0}^{M-1} h_{ik}^{(e)} \pi_k h_{kj}^{(d)} = \mathbf{r}^T \text{diag} (\mathbf{Y} \Theta^T),$$

where $\Theta = [h_{ij}] = \mathbf{r} \odot \text{diag}(H(f))$ and $\Theta = [\theta_{nj}] \in \mathbb{R}^{M \times M}$ is $\Theta = H(e) \Pi^{(d)} = \Pi^{(d)} H(e)$; $\theta_{nj} = \sum_{k=0}^{M-1} h_{jk}^{(e)} \sum_{m=0}^{M-1} \pi_{km} h_{mj}^{(d)} = \sum_{k=0}^{M-1} h_{jk}^{(e)} \pi_{kk} h_{kj}^{(d)}$. The operator $\text{diag} (\mathbf{A}) = [a_{ij}]$ extracts the vector of diagonal elements of matrix $\mathbf{A}$. 
The scalar normalized average goodput corresponds to the average goodput of a system adopting a hypothetical set of modes all having unitary transmission rate:

$$\bar{R}_c = \| \text{diag} (Y \Theta^T) \|_1. \quad (4)$$

The use of this quantity will be clear in the analysis.

Let us now see how the expression of the effective capacity specializes in few special cases. Under the assumption of additional blind mode detection, a mode acquisition error does not imply information loss but only mismatch in the used mode. In this case

$$\bar{R}_c = \sum_{i=0}^{M-1} r_i \sum_{j=0}^{M-1} (1 - e_{ij}) \sum_{h=0}^{M-1} h_{ih}^{(f)} \sum_{k=0}^{M-1} h_{kh}^{(c)} \pi_k h_{kj} = \sum_{i=0}^{M-1} r_i (1 - e_{ii}) \pi_{ii} = \text{tr} \text{diag} (Y \Pi), \quad (5)$$

where $\Theta = H^{(f)} H^{(c)} \Pi H^{(d)} = \Pi H^{(d)} H^{(f)} H^{(c)}$. In this case, $\Theta = \Pi \Omega$, where $\Omega$ is the equivocation matrix, which includes all the imperfections of link adaptation processes.

In the ideal case of perfect link adaptation, in which $M_i$ timely follows without errors $S_i$, $H^{(c)} = H^{(d)} = H^{(f)} = I$, and (3) reduces to

$$\bar{R}_c |_{\text{ideal}} = \sum_{i=0}^{M-1} r_i (1 - e_{ii}) \pi_{ii} = \text{tr} \text{diag} (Y \Pi). \quad (6)$$

Fixed systems are considered in this model as a special case with $H^{(f)} = I$, $H^{(d)} = I$, and static mode selection leading to a matrix $H^{(c)}$ having as non-zero elements all 1 in the $i$-th row, if mode $M_i$ is implemented in the system: $h_{hi}^{(c)} = \delta_{hi} \delta_{kk}$, where $\delta_{ij}$ is the Kronecker delta.

V. ILLUSTRATIVE EXAMPLES

For illustration purposes, we apply the above analysis tool to a sample system to investigate the sensitivity to imperfections and implementation constraints. The purpose is to show how imperfections influence the effective capacity and the importance to include those into the model. It is out of the scope of this paper to provide performance analysis of a specific system.

Assumptions. Consider a five-mode link-adaptive system with BPSK, QPSK, 8QAM, and 16QAM modulation schemes, and a no-transmission mode for insufficient signal quality. The same channel code ($g_{01} = 133s_8, g_1 = 171s_8, K = 7$) is used for all modes [1]. Service requirement on the PER as $e_{\text{max}} = 10^{-5}$ is assumed. With independent bit errors at link layer (achieved, e.g., with sufficiently long interleaver), the PER is $P_E = 1 - (1 - p_e^c (\gamma_{G_c} (\gamma)))^L_p$, where $p_e^c (\gamma)$ is the bit error rate of the uncoded system, $G_c (\gamma)$ is the coding gain, $\gamma$ is the SINR per bit, and $L_p$ is the packet length in bits. The assumption of independent bit errors and the use of $G_c (\gamma)$, used in this example, are not needed in the generic model of Sect. IV. Markov models are widely adopted in the literature for channel modeling, since they lead to tractable models. The validity of first-order Markov models is addressed, e.g., in [6] [7]. Continuous time models may integrate better with other fluid analytical models and are efficiently implemented in event-driven simulators. We model the presence in the $M$ signal quality regions $S$ with a continuous time Markov chain with infinitesimal generator $Q \in \mathbb{R}^{M \times M}$ [1]. The assumption of Markov chain is however not needed in the generic model of Sect. IV, since only the stationary state probabilities appear in the formula of the effective capacity.

![Fig. 1. Sensitivity of stationary state probabilities to hysteresis region width.](image)

For this illustrative example, no assumption is done on the structure of the estimator, and Gaussian distributed error is assumed. In this case, the elements of $H^{(c)}$ are $h_{kk}^{(c)} = 0.5 \{ \text{erfc} (\gamma_k - \gamma_{k})/\sigma_e \} - \text{erfc} (\gamma_{k+1} - \gamma_{k})/\sigma_e \}$, where $\text{erfc} (\cdot)$ is the Gaussian complementary error function, $\sigma_e^2$ is the estimation error variance, and $\gamma_{k}$ is the nominal value of the metric in $S_k$. The nominal value may be the average value in the region. With $\gamma_e = 0$ and error free acquisition channel, the sensitivity of $\bar{R}_c$ to estimate errors is studied. In Fig. 2, $\bar{R}_c$ exhibits a maximum for variance larger than zero. This behavior is explained by the fact that, in the expression for $\bar{R}_c$, the success rate is weighted by the bit rate. The scalar normalized average goodput $\bar{R}_c$, defined in 4, takes into
account only the effects of the error rate $e$ and is independent of the transmit bit rate $r$. As expected, $R_c$ is monotonically decreasing. The maximum of $R_c$ is due to a small gain in using a larger bit rate even if combined with a slightly larger error rate.

Sensitivity of Effective Capacity to Estimation Delay. To simplify the analysis, in this example the estimation delay is assumed negligible compared to the minimum inter-transition time (time between state changes): $\lambda_i \tau_e = q_i, i+1 \tau_e \ll 1$ and $\mu_i \tau_e = q_i, i-1 \tau_e \ll 1$. Under this assumption, it is straightforward to see that $H^{(d)} = I + Q \tau_c$. In this case, $R_c(\tau_c) = r^T \text{diag}(Y \Theta^T) = R_c(0) + \tau_c \Psi_d$, with $R_c(0)$ given by (6) and where $\Psi_d = r^T \text{diag}(YQ^T \Pi)$ is the unitary drift of the effective capacity from the ideal conditions due to the estimation delay.

Sensitivity of Effective Capacity to Acquisition Errors. To identify a mode, a control message with $m = \lceil \log_2 M \rceil$ bits is used. The distance in bits among all pairs of codewords is given by a symmetric matrix having null diagonal. $D = \{d_{ij}\} \in \mathbb{N}^{M \times M}, d_{ij} = \sum_{n=1}^{m} w_i^{(n)} \otimes w_j^{(n)}$, where $w_i^{(n)}$ is the $n$-th bit of the $i$-th codeword and $\otimes$ denotes the modulo 2 bitwise product. Mode identifiers may be coded with the Gray code. Assume that the message is transmitted always using the strongest mode. It has been seen that the effect of acquisition errors is negligible. In fact, as shown in Fig. 3, in which the bit error rate in the acquisition channel, $p_{e^{(f)}}$, is assumed fixed and independent of the channel state, the impact is no longer negligible only for $p_{e^{(f)}}$ out of useful range for adaptive and communications systems in general.

VI. CONCLUSIONS

In this paper, a model for the service capacity at link layer has been presented. At link layer, resource allocation protocols manage resources made available by the physical layer and assign those to services offered at their upper layers. A good metric for characterizing the available resources, and therefore the effective link service capacity, is the goodput. The presented effective capacity model incorporates a number of characteristics, including link adaptation with its imperfections. Illustrative examples showed that aspects above should not be neglected for realistic performance analysis at upper layers.

The effective capacity is expressed also in a compact form including the impact of imperfections and implementation losses. The characteristics of the adaptive system are independently represented by separate matrices, for the lower layers ($Q$, $\pi$, $r$, $y$) and for the imperfections in the transceivers set-up ($H^{(c)}$, $H^{(d)}$, $H^{(f)}$). Therefore, the model can be flexibly adapted to a number of systems. Consistent comparison of different scenarios can be done by using the same model and changing properly the model’s matrices. Mobility may be included in the model by averaging over the channel conditions represented by distinct matrices $Q$. The proposed model can also be incorporated into network simulators, implementing the separate look-up tables (matrices). In a multi-user scenario, distinct matrices for each user are used.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their comments, which improved the quality of the paper.

REFERENCES


