Analytic study of packet delay from 4G and 5G system ARQs using Signal Flow Graphs

Sebastian Lindner, Jon David Kroening, Phuong Nga Tran, Christoph Petersen, Andreas Timm-Giel
Institute of Communication Networks (ComNets)
Hamburg University of Technology (TUHH)
Hamburg, Germany
{sebastian.lindner, jon.david.kroening, phuong.tran, petersen, timm-giel}@tuhh.de

***Abstract*—Modern mobile networks in the era of 5G have ambitious performance goals. One envisioned goal is to have communication that is ultra-reliable and ultra-low latency (uRLLC). To ensure reliability, Automatic Repeat Request (ARQ) processes are usually deployed, sometimes combined with forward error correction. As an evolution of the Long Term Evolution (LTE) system, 5G may be foreseen to use similar methods, and so we assume LTE's Hybrid ARQ on the Medium Access Control (MAC) layer, and a Selective Repeat ARQ on the Radio Link Control (RLC) layer. Simulation of LTE systems is a time-consuming endeavor due to its complexity, and when latency and reliability are the key performance indicators, it may be preferable to isolate the ARQ processes. In this manner, this paper presents a mathematical model of the two stacked ARQ processes that are used in today's LTE system. It is based on the method of Signal Flow Graphs (SFGs) and allows the analysis of either ARQ process in isolation, and of both stacked on top of each other. Analytic results are shown for a realistic set of parameters, and in the spirit of Open Science, the MATLAB implementation files are made available in the conclusion to interested researchers, so that verification and an adaption to similar ARQ processes is easily made.**

***Index Terms*—LTE systems, 4G mobile systems, 5G mobile systems, automatic repeat request, signal flow graphs**

I. Introduction

Fifth Generation (5G) mobile systems promise an application-specific set of system characteristics, so that vastly different use cases are supported. To enable the operation of novel functions such as vehicle-to-infrastructure (V2I) and autonomous driving, according to the ITU-R report in [1], ultra-reliable low latency communication (URLLC) is required with a latency in the range of 1 ms and reliability of less than one packet loss in 10^5 packet transmissions.

To study whether a mobile system achieves the required key performance indicators (KPIs), often-times simulations and emulations are performed. With systems as complex as Fourth Generation (4G) LTE or 5G systems, such simulations are time-consuming, and emulation is difficult. An analytic approach promises much faster results, and allows the isolation of the sub-system in question; in this case, the ARQ processes used in 4G and 5G systems. SFGs are a powerful analytical tool that is capable of modeling complex communication systems statistically.

A SFG models the propagation of a *signal* traversing the graph nodes. Channel burst errors that are correlated in time can be modeled using a Hidden Markov Model (HMM). As the LTE system uses discrete time slots, the discrete-time Probability Generating Function (PGF) is then read off the graph after simplification. The PGF holds all information of the probability distribution of the statistic of interest, and so full knowledge about the distribution of packet delays can be acquired.

The paper is organized as follows. The history of SFGs applied to communication networks is briefly presented in Section II. The system model, the mathematical basics and notation of SFGs and relevant components of the LTE system are described in Section III. Section IV explains the SFGs that model these components. In Section V the results are evaluated, and Section VI gives a conclusion.

II. Related work

In 1989, SFGs were used by the authors of [2] to study ARQ processes with erroneous feedback channels. In particular, the Go-back-N ARQ and Selective Repeat ARQ algorithms were investigated.

The authors of [3] study a Selective Repeat ARQ process, where both the forward and reverse channels are modeled using HMMs. In the paper, the idea of incorporating a HMM into the Matrix Signal Flow Graph (MSFG) is introduced, so that time-correlated burst errors can be modeled. They employ "simulation" to verify their results, but no details are given into how these simulations are configured or implemented.

The paper in [4] uses MSFGs to model a Selective Repeat ARQ process. It specifically addresses general Selective Repeat ARQ processes, as the analysis is valid for all systems that employ these. The selective retransmission of only those packets that were lost is modeled through recursion in the MSFG.

In [5] the throughput and average delay of Hybrid Automatic Repeat Request (HARQ) is compared to a Selective Repeat ARQ, where erroneous feedback is considered through MSFGs that contain a Gilbert-Elliot HMM. The loss reduction through Soft Combining in HARQ is considered, which means that a performance improvement through network coding is part of the process. In comparison to this work, the misinterpretation of a negative acknowledgment (NACK) as an acknowledgement (ACK) is not considered. When this misinterpretation happens, then the sender wrongly assumes a successful transmission has taken place and deletes the current packet from the retransmission buffer.

The authors of [6] study the capacity and error ratio of the physical, MAC and RLC layers of the LTE protocol stack. A fixed block error rate and an erroneous feedback channel are considered, and the focus is put on HARQ on the MAC, and on segmentation and Selective Repeat ARQ on the RLC layer. As for HARQ, no Forward Error Correction (FEC) is considered, and thus the ARQ is not really *hybrid*.

In [7] the authors discuss possible ARQ variants for 5G systems. The general architecture of both MAC- and RLC-layer ARQs stays valid as it was for 4G systems, and so our analysis focuses on this setup.

In comparison to the related work, this paper presents an analytic approach to the analysis of LTE's HARQ that include the possibility of misinterpreting ACKs (other authors consider only their loss, which leads to an increased delay, while a wrongly interpreted NACK leads to the transmission attempt failing completely); and of the RLC layer ARQ process that is based on the findings of the MAC layer HARQ SFG. Also, our results are compared to a recent simulation framework, which other works usually lack.

III. System Model

In LTE, typically communication takes place between a User Equipment (UE) and the evolved NodeB (eNB). Two different variants of ARQ are employed by the eNB to ensure reliable communication: 1) RLC layer packets are protected through a Selective Repeat ARQ, 2) MAC layer segments are protected by up to 8 parallel HARQ processes. Each HARQ process is a Stop-and-Wait ARQ; according to [8]. The process is *hybrid* as it combines ARQ with FEC, eliminating the need of some retransmissions. Additionally, Soft Combining is a technique where the receiver buffers erroneous packets and uses them as additional, redundant information to decode later packets. These assumptions are true for both 4G and 5G systems. As it is difficult to obtain realistic parameters for 5G systems, our analysis focuses on known values for 4G, but the models are general enough to be easily extended to 5G parameters as they become available.

A SFG is a directed graph $G = (N, E)$ where an incoming edge depicts a dependency on the previous node. A signal travels from the input node via internal nodes to the output node; on the way it is augmented on each edge (i, j) by a function G_{ij}. The signal at node `k` – here, node notation is in typewriter font – is $n_k = \sum_{j=1}^{|N|-1} n_j \; G_{jk}, \; \forall k$. A SFG can be simplified according to the series, parallel, self-loop and splitting rules, whose thorough descriptions can be found in [9]. The packet delay between sender and receiver corresponds to a random variable X whose behavior should be studied. After simplifying a SFG, the random variable's PGF can be read off, which resides in the z-domain, a discrete-time equivalent of the Laplace-domain. Generally a PGF is defined in Eq. 1.

$$G_X(z) = \mathbb{E}[z^X] = \sum_{x=0}^{\infty} z^x \cdot P(X = x) \tag{1}$$

The Probability Mass Function (PMF) $P(X = x)$ is a component of Eq. 1, and can be recovered through Eq. 2.

$$P(X = k) = \frac{1}{k!} \frac{d^k}{dz^k} G_X(0), \quad k = 0, 1, 2, \ldots \tag{2}$$

Figure 1 shows a discrete-time Gilbert-Elliot channel model where P_A corresponds to the channel being in a good state `A` with a small error probability ϵ_A, and P_B corresponds to the channel being in a bad state `B` with a large error probability ϵ_B.

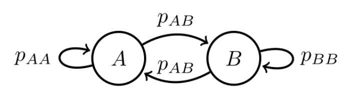

Fig. 1: A Gilbert-Elliot model for correlated losses.

The transition and error probabilities are captured in $P = \begin{bmatrix} 1 - p_{AB} & p_{AB} \\ p_{BA} & 1 - p_{BA} \end{bmatrix}$ and $L = \begin{bmatrix} \epsilon_A & 0 \\ 0 & \epsilon_B \end{bmatrix}$. The stationary distribution $\pi = [\pi_A \quad \pi_B]$ of the model is found as

$$\begin{aligned} \pi \cdot P &= \pi \\ \pi \cdot \mathbf{1} &= 1 \\ \Rightarrow \pi &= \begin{bmatrix} \frac{p_{BA}}{p_{AB} + p_{BA}} & \frac{p_{AB}}{p_{AB} + p_{BA}} \end{bmatrix} \end{aligned} \tag{3}$$

IV. Graphs

A. Stop-and-Wait (SW)-ARQ with erroneous feedback

The number of times a packet may be transmitted by a SW-ARQ is limited to $N = 5$ in LTE's HARQ according to [10]. With this, a SW-ARQ can be modeled as in Fig. 2, where the states correspond to the number of retransmissions so far. For every transmission attempt up to N, the transmission can succeed, and the signal travels to the success state `S`. At state `N`, the final transmission might fail, so the failure state `F` can be reached.

v corresponds to the Round Trip Time (RTT) that passes per transmission, success is modeled through $(I - L)$ with I being the identity matrix and the channel conditions are evaluated every time slot as P.

When feedback can be erroneous, in LTE this translates to ACKs being interpreted as NACKs and vice-versa. The third possible response of a discontinuous reception (DTX) leads to the same reaction as a NACK; thus here they are viewed as synonymous. The probability matrix $E_{NA} = \begin{bmatrix} e_{NA,A} & 0 \\ 0 & e_{NA,B} \end{bmatrix}$ represents this misunderstanding, where the probability $e_{NA,B}$ corresponds to a NACK being interpreted as an ACK when the channel is in state B. $E_{NN} = I - E_{NA}$ is thus the probability matrix for correct interpretation.

In Fig. 2, state `i` corresponds to i transmission attempts, and intermediate state `i'` to the decision state, where according to d the NACK is correctly interpreted, and according to c it is

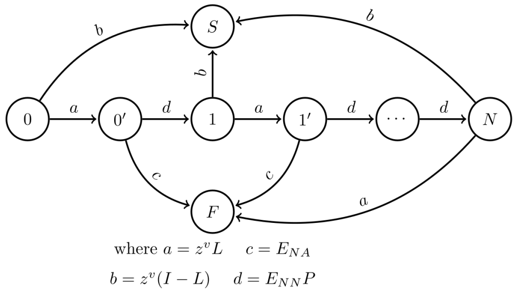

Fig. 2: MAC-Layer SW-ARQ MSFG with at most N retransmissions and an erroneous feedback channel.

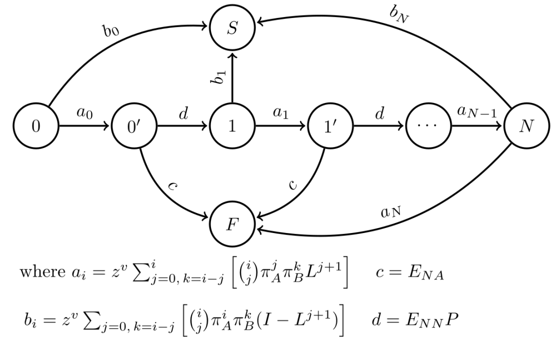

Fig. 3: MAC-Layer HARQ MSFG with at most N retransmissions and an erroneous feedback channel.

not. A special case is the final transmission state N, where the failed state is reached upon a loss, no matter how the NACK is interpreted. The transfer functions are derived by simplifying the SFG in Fig. 2 using the aforementioned rules into an atomic graph from which they can be read off. They are given in Eq. 4 and the full PGF is their sum in Eq. 5.

$$\begin{aligned} T_{0\to S}^{\text{SW-ARQ-err}} &= \sum_{i=0}^{N} (z^v L P E_{NN})^i z^v (I - L) \\ T_{0\to F}^{\text{SW-ARQ-err}} &= z^v L \sum_{i=0}^{N-1} (P E_{NN} z^v L)^i E_{NA} \\ &\quad + (z^v L P E_{NN})^N z^v L \end{aligned} \tag{4}$$

$$G^{\text{SW-ARQ-err}} = T_{0\to S}^{\text{SW-ARQ-err}} + T_{0\to F}^{\text{SW-ARQ-err}} \tag{5}$$

B. HARQ with Soft Combining

With Soft Combining, the coding information of received, erroneous packets is used to decode later, also erroneous packets. Insofar the error probability decreases as more retransmissions occur. In Fig. 3 the transitions to the success state become state-specific with transfer functions b_i. In state i, there are $\binom{i}{j}$ combinations of observing the good channel state A j times, and so the sum goes over all combinations of good and bad channel state observations. Without loss of generality, an exponential decrease of the packet error probability is assumed with each retransmission.

Figure 3 leads to the transfer functions in Eq. 6 and the PGF in Eq. 7.

$$\begin{aligned} T_{0\to S}^{\text{HARQ}} &= \sum_{i=0}^{N} \prod_{j=0}^{i-1} [a_j d] b_i \\ T_{0\to F}^{\text{HARQ}} &= a_0 \sum_{i=0}^{N-1} \prod_{j=1}^{i} [a_j d] c_1 + \prod_{j=0}^{N-1} [a_j d] a_N \end{aligned} \tag{6}$$

$$G^{\text{HARQ}}(z) = T_{0\to S}^{\text{harq}} + T_{0\to F}^{\text{harq}} \tag{7}$$

C. Radio Link Control layer ARQ

In the LTE protocol stack, the RLC layer sits above the MAC layer. Here, a Selective Repeat ARQ protocol may coordinate the retransmission of RLC Protocol Data Units (PDUs). When enabled, *status reports* sent by the receiver indicate which RLC PDUs have been lost, so that these can be selectively retransmitted. According to [11], such status reports can be requested by the sender, or sent autonomously upon packet loss.

Here, the focus is put on the receiver-sent status report, when an out-of-order reception has taken place. The sender initiates a timer after being notified of an out-of-order reception, which expires after the last moment in time when the PDU may still arrive – it could be being retransmitted by an underlying HARQ instance and arrive out-of-order. When the timer expires after the N_{HARQ} retransmission attempts of the HARQ process, a retransmission in the RLC layer is initiated.

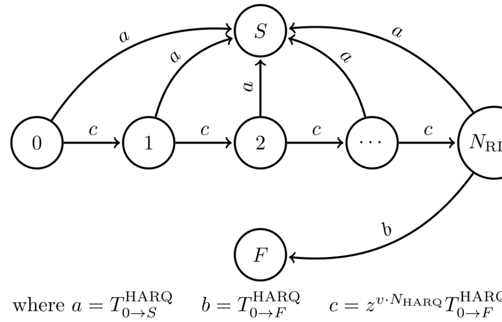

Fig. 4: RLC-Layer ARQ MSFG with at most N_{RLC} retransmissions and an underlying HARQ process.

Fig. 4 depicts the discussed RLC layer ARQ. The nodes again denote the number of retransmissions attempted so far, where one retransmission refers to an entire run through a MAC layer HARQ process. Each branch that ends in the success state therefore holds the HARQ transition function $T_{0\to F}^{\text{HARQ}}$. When the HARQ process fails, the discussed timer has to expire before another HARQ attempt is initiated by the RLC layer; thus $z^{v \cdot N_{\text{HARQ}}}$ time steps have to pass – note that this is a best-case assumption where exactly when all retransmissions of the HARQ process are exhausted, the sender is assumed to know of the failed transmission. This is the first moment in time that the transmission has surely failed; before this moment, another retransmission could succeed. However, realistically the timer may not start because no out-of-order reception has taken place – think channel burst errors – or the timer starts later than this amount of time.

When the $\mathrm{N_{RLC}}$ state is entered, the transmission can finally succeed through the $T^{\text{HARQ}}_{0\to S}$-branch, or ultimately fail through the $T^{\text{HARQ}}_{0\to F}$-branch. This leads to the transfer functions in Eq. 8 and the PGF in Eq. 9.

$$\begin{aligned} T^{\text{RLC}}_{0\to S} &= \sum_{i=0}^{N} (z^{v \cdot N_{\text{HARQ}}} T^{\text{HARQ}}_{0\to F})^i T^{\text{HARQ}}_{0\to S} \\ T^{\text{RLC}}_{0\to F} &= (z^{v \cdot N_{\text{HARQ}}} T^{\text{HARQ}}_{0\to F})^N T^{\text{HARQ}}_{0\to F} \end{aligned} \tag{8}$$

$$G^{\text{RLC}}(z) = T^{\text{RLC}}_{0\to S} + T^{\text{RLC}}_{0\to F} \tag{9}$$

V. Results

The PGFs in Eqs. 5, 7 and 9 as well as an ideal feedback channel case are evaluated in MATLAB. They are transformed to scalar form by left-multiplying with the steady state distribution π from Eq. 3 and right-multiplying with a vector of ones as in Eq. 10.

$$G_{(\text{scalar})}(z) = \frac{\pi G_{(\text{matrix})}(z)\mathbb{1}}{\pi \mathbb{1}} \tag{10}$$

Then the PMF is extracted as in Eq. 2. The probability $P(X = k)$ follows, where X is the random variable that represents the number of timeslots required for a successful packet transmission, and k its realization. The number of max. transmission attempts for one HARQ process is set to LTE's default value of $N_{\text{HARQ}} = 4$ and to $N_{\text{RLC}} = 3$ on the RLC layer. W.l.o.g. $p_{AB} = 0.25$, $p_{BA} = 0.5$ are chosen for the Gilbert-Elliot channel model, while the loss probabilities are $\epsilon_A = 0.1$ (LTE's target value achieved through adaptive modulation and coding), $\epsilon_B = 1$ (erroneous channel). Then the misinterpretation probability matrix E_{NA} is chosen to reflect LTE's target of $0.01\,\%$ in both channel states.

The verification of mathematical soundness is done through a simple check: with an upper-bounded max. transmission attempts, the evaluation over all time slots *must* result in a PMF that sums to one. The verification of systematic soundness is achieved through the comparison with the OMNeT++ simulator, available in [12], and the simuLTE framework that provides the LTE protocol stack, available in [13]. A single UE and a single eNB are simulated. $40\,\mathrm{B}$ packets are sent in intervals of $200\,\mathrm{ms}$ to the UE, which are small enough to be transmitted without segmentation, and infrequently enough to ensure that each packet is handled isolated from the next, because the transmission concludes before the next transmission begins: the same $N_{\text{harq}} = 4$ is configured, and a transmission takes $4\,\mathrm{ms}$. Simulations run for $10\,001\,\mathrm{s}$ and are repeated 50 times; $95\,\%$ confidence intervals are given (but are barely visible, indicating a large confidence).

Fig. 5 shows the CCDFs of the respective $T_{0\to S}$ transfer functions. LTE has a transmission time interval (TTI) of $1\,\mathrm{ms}$ and one transmission requires a RTT of $4\,\mathrm{ms}$, so the evaluation is done for $k = 0, 1, 2, \ldots, \text{RTT} \cdot (N_{\text{HARQ}} + 1) = 20\,\mathrm{ms}$.

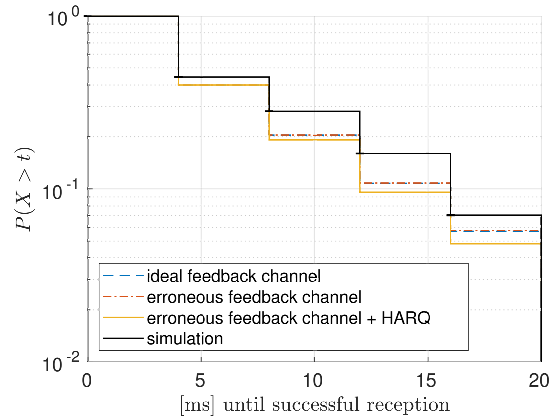

Fig. 5: Comparison of the Complementary Cumulative Distribution Functions (CCDFs) of MAC layer ARQs with ideal and erroneous feedback channels, as well as the hybrid HARQ approach through SFGs, and through simulation.

Ideal and erroneous feedback channel CCDFs diverge slightly for more retransmissions, as errors through false ACK/NACK identification occur. Since this likelihood is tiny in LTE, the probability of success is mostly identical. The integration of HARQ, however, shows a large benefit. While for the first transmission no benefit can be exploited, partial receptions of packets accumulate over time, and the next transmission attempt has a significantly larger chance of success – non-HARQ protocols are clearly outperformed.

Simulations follow a similar behaviour, but deviate for an increasing number of retransmissions. Note that the exponential decrease in the probability of an erroneous reception through HARQ's FEC is best visible on a logarithmic scale; the absolute y-deviation between the analytic SFG approach and the simulation-based approach is at most $8.93\,\%$ and on average $4.07\,\%$. As for the relative difference, on average the SFG model lies $64\,\%$ below the simulation results. This is accredited to a more detailed model of the LTE protocol stack in simuLTE, where effects may occur that are not present in the mathematical model, which views the ARQ in isolation from the rest of the protocol stack. We conclude that the simulation results verify the SFG models, and identify them as a lower bound to a realistic CCDF of packet delays. To our knowledge this is the first comparison of SFGs to simulation data for LTE ARQ mechanisms.

For the RLC layer ARQ, simuLTE has not implemented the respective protocol, so no comparison to simulation data exceeding the first HARQ process can be obtained. In the SFG model, the comparison over a longer time frame for the same configuration as before shows how for each RLC transmission attempt, the timer has to expire (a long plateau), and then the typical HARQ process "steps" can be seen in Fig. 6.

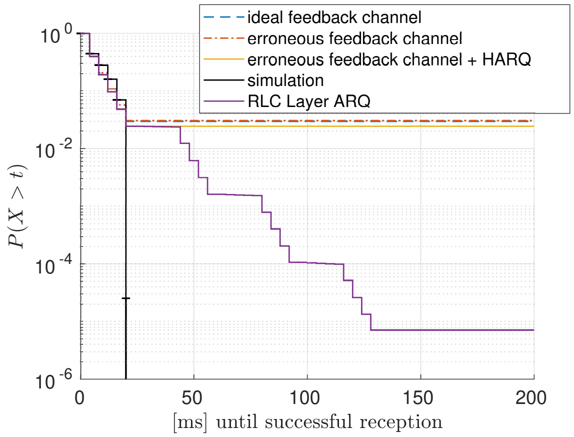

Fig. 6: Comparison of the CCDFs of MAC layer ARQs with ideal and erroneous feedback channels, as well as the hybrid HARQ approach through SFGs to the RLC layer ARQ mechanism.

VI. Conclusion

In this paper we have investigated the mechanisms of LTE's ARQ processes. On the MAC layer, the Hybrid Automatic Repeat Request process incorporates Forward Error Correction so that future transmission attempts benefit from an increased chance of successful reception. Our comparison between Stop-and-Wait-ARQs in an ideal feedback channel, in an erroneous feedback channel, and a HARQ in an erroneous feedback channel have shown how the FEC improves system performance considerably. But most importantly, we have developed a Signal Flow Graph model for this type of ARQ that allows the analytic investigation of its behaviour. Next a joint view on RLC and MAC layer ARQs is considered. We have designed a new RLC layer ARQ model that makes use of the MAC layer transfer functions of the first analysis, instead of an isolated view on the two ARQs as usually done. With this, a mathematical model has been developed that covers both layer ARQs of the LTE communication system jointly, and that considers the technical details of this system in considerable depth: FEC eases the decoding of successive retransmissions, an erroneous feedback channel may lead to the incorrect interpretation of ACKs / NACKs, and the RLC layer timer has an impact.

An identification of currently unmodeled components of the LTE protocol stack, which influence the packet delay, should be attempted in order to close the gap between simulation and analytic results. Currently, our approach yields a lower bound for the CCDF of packet delays, as the full LTE protocol stack influences this metric, and our model views solely the ARQ process. An interesting outlook would be a more in-depth investigation of 5G system parameters as they become available. Due to limited access to realistic values, we so far focused on 4G parameters. Another aspect is the integration of multi-hop Device-to-Device (D2D) communication, envisioned to become of major significance in 5G systems. If each hop functions in the same way as a here-depicted eNB, a multiplication of the transfer functions with themselves, equivalent to a convolution in the time domain, can be done; otherwise, the functional differences have to be modeled. MATLAB code and the simulation model can be found in [14], while the custom branch of the simuLTe framework can be found in [15].

References

[1] ITU-R. *ITU-R M.2410.0: Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)*. 2017.

[2] D.-L. Lu and J.-L. Chang. "Analysis of ARQ Protocols via Signal Flow Graphs". In: *IEEE Transactions on Communications*. Vol. 37. IEEE, 1989, pp. 245–251.

[3] Kamtorn Ausavapattanakun and Aria Nosratinia. "Analysis of Selective-Repeat ARQ via Matrix Signal-Flow Graphs". In: *IEEE Transactions on Communications*. Vol. 55. 2007.

[4] Guido Gehlen and Bernhard Walke. "Transport Layer Delay Analysis in Wireless Networks Using Signal Flow Graphs". In: *Proceedings of European Wireless 2007*. 2007.

[5] Derya Malak, Muriel Médard, and Edmund M.Yeh. "Analysis of Coded Selective-Repeat ARQ via Matrix Signal-Flow Graphs". In: *arXiv Preprint arXiv:1801.10500*. 2018.

[6] Yuan Chen and Bernhard Walke. "Analysis of Capacity and Error Ratio in 3GPP LTE Systems Using Signal Flow Graph Models". In: *European Wireless 2013; 19th European Wireless Conference*. VDE, 2013, pp. 1–8.

[7] Cong Shi et al. "Retransmission Schemes for 5G Radio Interface". In: *2016 IEEE Globecom Workshops (GC Wkshps)*. 2016 IEEE Globecom Workshops (GC Wkshps). Dec. 2016, pp. 1–6. DOI: 10.1109/GLOCOMW.2016.7849030.

[8] A Larmo et al. "The LTE Link-Layer Design - [LTE Part II: 3GPP Release 8]". In: *Communications Magazine, IEEE* 47.4 (2009), pp. 52–59. DOI: 10.1109/MCOM.2009.4907407.

[9] Ronald A. Howard. *Dynamic Probabilistic Systems: Markov Models*. Vol. 1. Courier Corporation, 2012.

[10] 3GPP. *Universal Mobile Telecommunications System (E-UTRA); Radio Resource Control (RRC)*. 3GPP, 2018.

[11] Matthew Baker Stefania Sesia Issam Toufik. *LTE - The UMTS Long Term Evolution: From Theory to Practice, 2nd Edition*. Wiley, 2011. ISBN: 978-0-470-66025-6.

[12] *OMNeT++*. URL: https://omnetpp.org/ (visited on 09/12/2017).

[13] *simuLTE*. URL: http://simulte.com/ (visited on 09/12/2017).

[14] Sebastian Lindner. *MATLAB Code and Simulation Model*. 2020. URL: https://doi.org/10.5281/zenodo.3691880.

[15] Sebastian Lindner and Raphael Elsner. *Custom simuLTE Framework with Gilbert-Elliot Channel Model Implementation*. 2019. URL: https://doi.org/10.5281/zenodo.3351385.