Feature Extraction and Classification for Communication Channels in Wireless Mechatronic Systems

Jing Geng*, Mohamed Kashef†, Richard Candell†, Yongkang Liu†, Karl Montgomery†, Shuvra S. Bhattacharyya*

* Department of Electrical and Computer Engineering, University of Maryland, College Park, USA

† Intelligent Systems Division, Engineering Laboratory, National Institute of Standards and Technology, USA

Email: jgeng@umd.edu, {mohamed.kashef, richard.candell, yongkang.liu, karl.montgomery, }@nist.gov, ssb@umd.edu

***Abstract*— For accurate characterization and evaluation of wireless mechatronic systems, effective modeling of wireless communication channels is of paramount importance, especially to simulation-oriented methods. Conventional simulation methods employ mathematical models to abstract details of prototype channels. Although such mathematical models often have rigorous theoretical underpinnings, they can be weak in capturing complex environmental characteristics and complex forms of diversity that are exhibited in industrial communication environments. To address this problem, we develop, in this paper, a new approach to deriving effective simulation models for industrial communication channels. Our approach involves field measurements from actual wireless mechatronic environments together with feature extraction from the measurements, and data-driven classification based on the extracted features. Our approach leads to a general framework for simulating wireless mechatronic systems in a way that realistically incorporates the complex channel characteristics of these systems.**

I. Introduction

In recent years, integrating wireless communications capabilities into industrial mechatronic systems has attracted great interest due to promising advantages offered by wireless communications [1]. This integration brings about highly complex design spaces, which we refer to as *wireless-integrated factory system* (*WIFS*) design spaces, involving interactions among physical factory layout, control algorithms, and wireless communication networks (e.g., see [2]). General and effective simulation methods are needed for WIFS environments and other types of complex mechatronic environments for performance assessment of existing system designs, development of new designs, and planning of technology updates.

Many software tools have been developed to provide sophisticated capabilities for communication network simulation. However, traditional radio frequency (RF) wireless channel models that are used with such tools have significant limitations in the context of simulating WIFS environments. Such models typically utilize mathematical formulations associated with collected data or power delay profiles (PDPs) obtained from third-party studies (see Section II). WIFS simulations with such models may not accurately reflect actual system performance since the models do not take into account harsh conditions and other distinguishing characteristics of RF channels in industrial communication environments. Such characteristics arise, for example, from vibration of machinery, and the presence of metal objects (e.g., see [3]), which can have a major impact on communication performance.

In this paper, we build upon our recent work that introduced a new link layer simulation approach for integrating field measurements into WIFS simulation [4]. We refer to that method as the Measurement-based Channel Library Generator (MCLG) method, which involves constructing channel library modules, in the form of PER/SNR (packet error rate / signal-to-noise ratio) tables, from field measurements in a systematic manner. However, the MCLG method has a couple of limitations. First, the clustering algorithm used based on [5] is computationally intensive, and the running time can become very large with large measurement datasets. Second, the derived channel libraries correspond specifically to the measured environment for the specific measured communication paths. In large source environments, it is unrealistic to measure channel impulse responses (CIRs) for all possible communication paths.

In this paper, we develop significant improvements to the MCLG method that address the two limitations described above, and allow the method to be used more generally, in a broader class of WIFS design space exploration scenarios for a given source environment. We refer to the improved version of the MCLG method as the Generalized Measurement-based channel Library Generator (GMLG) method, and we refer to our prototype implementation of the method as GMLG.

II. Related Work

Candell et al. discuss challenges involved in employing industrial wireless technology in mechatronic systems, and present guidelines for addressing those challenges [6]. A large body of literature addresses the modeling of wireless communication channels with emphasis on different communication modeling concerns (e.g., see [7], [8]).

Various works have investigated the modeling of communication channels in more challenging communication environments. For example, Abbas et al. performed comparisons between simulations and measurements to evaluate vehicle-to-vehicle communication channel parameters [9]. Peil et al. developed channel models using wireless propagation characteristics in an industrial environment [10]. Other works emphasize application of existing simulation frameworks to design challenges in specific application areas (e.g., see [2], [11]–[13]).

The GMLG method presented in this paper differs from previous works, including those summarized above, in its emphasis on systematically incorporating field measurements into simulation of wireless mechatronic systems, and its application of feature clustering to generalize the types of communication channels that can be simulated using a given set of field measurements.

III. Methods

Fig. 1 provides a block diagram illustration of the GMLG method. Input to GMLG consists of a set $S_c = \{C_1, C_2, \ldots, C_m\}$ of CIR measurements from distinct physical communication paths. The output includes a set $S_k = \{K_1, K_2, \ldots, K_n\}$ of representative channels along with a PER/SNR table $T(K_i)$ that characterizes the communication conditions represented by each K_i. The output also includes a machine learning model M that maps features extracted from communication paths in the source environment into representative channels. The primary components of MCLG that are reused in GMLG are the Link Level Simulator and Pre-processing block (see Fig. 1). A new clustering process is devised in GMLG to be more computationally efficient and to support the objective of deriving the classification model M.

In summary, the GMLG method processes CIR measurements to produce as output a triple (S_k, T, M), where T provides a mapping of the representatives in S_k into corresponding PER/SNR tables, and M provides a mapping of features extracted from arbitrary communication paths into S_k. Each ordered pair $(K_i, T(K_i))$ is referred to as a *library module* that is generated by the $GMLG$ method, and can be plugged into system-level simulators to aid in exploring WIFS design spaces.

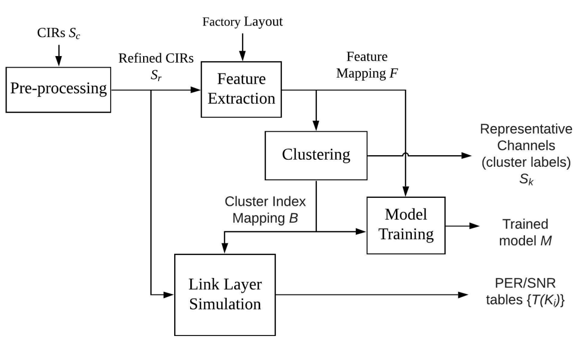

Fig. 1. An illustration of the GMLG method.

A. Feature Selection and Extraction

The input to the Feature Extraction block in Fig. 1 is a set of *refined* CIRs $S_r = \{R_1, R_2, \ldots, R_n\}$ that is produced by the Pre-processing block. Operations performed in the Pre-processing block include noise filtering, intra-CIR compression, and peak power determination, which are discussed in [4]. The output of the Feature Extraction block is a feature mapping $F : S_r \to S_f$, where for each $R_i \in S_r$, $F(R_i)$ is the feature vector extracted from the refined CIR R_i.

At this point, it is useful to distinguish between two types of features that are employed in GMLG — *external* and *system* features. External features are those that are input to the generated machine learning model M when applying M to higher level simulations by users. It should be possible to derive external features conveniently from data associated with the layout of the source environment. That way, new communication paths can be predicted by the model M even though they are not included in the measurements. On the other hand, system features are the features that are used to train the model M. In general, the system feature set includes all of the external features along with zero or more additional features. System features can be defined in terms of field measurement data — for example, they can be extracted from the CIRs collected to form the feature vector S_f. In GMLG, we have incorporated three external features, which are the path distance, line of sight (LOS) / non-LOS (NLOS) indicator, and Rician K Factor, with two additional features — mean delay and root mean square (RMS) delay spread — that are included in the system feature set.

B. Clustering

The Clustering block in Fig. 1 partitions refined CIRs into groups of related CIRs. Each group or *cluster* corresponds to a distinct channel library module that is generated by GMLG. A distinguishing aspect of GMLG is that clustering is performed in the feature space — that is, the clustering algorithm operates on the feature vectors rather than on the refined CIRs themselves. This enables much more efficient clustering since the feature vectors are highly compressed representations of the refined CIRs.

A wide variety of clustering techniques can be used in the GMLG method. Different clustering algorithms may be chosen depending, for example, on the spread of data points for each feature. In GMLG, we apply spectral clustering [14], which is known for its effectiveness in handling categorical data, and for its ability to handle inseparable data and derive non-convex clusters.

C. Clustering Assessment and Weight Optimization

For assessment of clustering solutions, we consider two different metrics, the silhouette coefficient (e.g., see [15]), and the average feature variance. In our context, a cluster corresponds to a set of refined CIRs that are grouped together by the Clustering block (Fig. 1), and a clustering solution corresponds to a partitioning of all refined CIRs into disjoint subsets (candidate clusters).

To assess the quality of clustering solutions in GMLG, we employ a composite metric that is formed of different sub-metrics, which are listed below. For a given sub-metric, we refer to the feature set used in assessment as the *evaluation feature set*, which is a subset of the five system features.

- Γ_1 is defined as the silhouette coefficient using an evaluation feature set that consists only of the K factor, LOS/NLOS indicator, and path distance (external feature set of GMLG).
- Γ_2 is defined as the silhouette coefficient using an evaluation feature set that consists only of the K factor, mean delay, and RMS delay spread. We refer to these features as ***quality assessment features*** since they are representative features for capturing channel characteristics from a measured CIR. The sub-metric Γ_2 is the main sub-metric for guiding the tuning process of clustering in GMLG.
- $\Gamma_3, \Gamma_4, \Gamma_5$ are variance metrics for the quality assessment features — K factor, mean delay, and RMS delay — respectively.

Although the external feature set is used for feature clustering, all five metrics $\Gamma_1, \Gamma_2, \ldots, \Gamma_5$ are used to guide the tuning process for feature weights when invoking spectral clustering. In other words, GMLG executes spectral clustering iteratively using different relative weightings of the features. Intuitively, the output solution B is selected as a solution that maximizes Γ_2, while retaining reasonable performance on the other four metrics.

D. Model Training

The clustering solution provided by the Clustering block is used to train a machine learning model, as illustrated by the Model Training Block in Fig. 1. The objective is to derive a trained machine learning model M that can map novel paths, based on estimated features from those paths, into representative channels.

The input to Model Training are the feature vectors $F(R_1), F(R_2), \ldots, F(R_N)$ for the refined CIRs along with the corresponding cluster indices $B(R_1), B(R_2), \ldots, B(R_N)$. The cluster indices are used as labels for supervised learning based on feature vectors. Intuitively, the model is trained to predict a well matched representative channel from a given feature vector composed of the external features.

There are many types of classifiers that can be applied within the GMLG method for the derivation of M. In GMLG, we apply the k-nearest-neighbors (KNN) algorithm [16], which is capable of handling classification functions involving categorical output results.

IV. Experiments

The set of field measurements that we use in our experiments consists of CIRs obtained from a measurement campaign from an automotive factory site performed by National Institute of Standards and Technology (NIST) [17]. We implement clustering and model training in Python (Version 3.7.4) and make use of the `scikit-learn` package [18] (Version 0.22).

A. Feature Clustering Results

The automotive factory site dataset includes $41{,}700$ CIRs, which are later reduced to $1{,}524$ refined CIRs after pre-processing. The results of the clustering process in GMLG on the automotive factory dataset are illustrated in Fig. 2. The figure provides a perspective on how clusters are formed in relation to different pairs of features. Each data point in each of the three plots corresponds to a feature vector that is projected onto the two-dimensional subspace corresponding to each plot. The data points are colored differently based on which cluster they are assigned to. The results illustrate that the clustering process is effective in separating the feature vectors into distinct regions of the feature space — this is perhaps most strongly demonstrated by the plot involving the K factor and path distance.

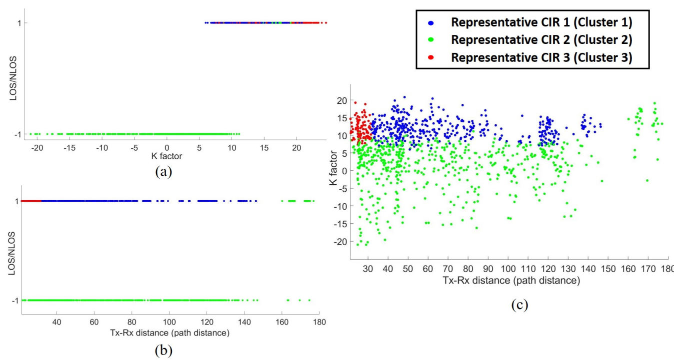

Fig. 2. Separation of feature vectors into clusters for the automotive factory dataset: (a) K factor vs. LOS/NLOS indicator, (b) path distance vs. LOS/NLOS indicator, (c) path distance vs. K factor.

B. Comparison Between AP-DTW and Feature Clustering

In this section, we present results that demonstrate significant improvements provided by the feature-based clustering method in GMLG compared to the clustering approach in MCLG, which is based on affinity propagation (AP) and dynamic time warping (DTW).

Because the computation time for the AP+DTW approach grows very rapidly with the dataset size, we perform the comparison experiment in this section using a smaller dataset with 254 processed-CIRs. This dataset is derived by using a larger downsampling factor during pre-processing.

The improvement in computational speed is significant. We found that the APT+DTW method required an average of $4{,}217$ seconds with standard deviation $\sigma = 194.0$, while GMLG required an average of 23.47 seconds with $\sigma = 5.64$. These results represent a speedup of 179.7X. The execution time measurements were averaged over 20 runs.

Table I compares the quality of the derived clustering solutions between AP+DTW and GMLG. The results show significant improvements in terms of all of the five evaluation metrics $\{\Gamma_i\}$ that were defined in Section III-C, especially for the silhouette metric Γ_1 and the variance metric Γ_3.

C. Classifying Novel Paths

As motivated in Section I, an important capability of the GMLG method is the ability to classify new communication paths (i.e., paths that do not correspond to any of the paths covered in the field measurements that are used to construct

TABLE I
COMPARISON OF CLUSTERING SOLUTIONS.

	AP-DTW Algorithm	Spectral Feature Clustering
Γ_1(silhouette coeff. based on external features)	0.091	0.571
Γ_2(silhouette coeff. based on quality assessment features)	0.151	0.249
Γ_3(feature variance for K factor)	40.15	26.49
Γ_4(feature variance for mean delay)	2336.84	2281.13
Γ_5(feature variance for RMS delay)	1213.36	979.98

the clusters). We evaluate this capability in our experiments by applying the derived classification model M on a *testing dataset* that is extracted from the original automotive factory measurement dataset, and that is excluded from the set of CIRs that is used in clustering and model training in GMLG. The paths in the testing dataset can be viewed as novel communication paths.

In this experiment, the classification labels produced in the experiment presented in Section IV-A are used as the ground truth. We then evaluate how accurately the model M, which is produced by GMLG, predicts the cluster label from the highly compressed (feature-based) representation that M operates on. We use 80% of the dataset as input to GMLG, which is configured to generate three representative channels. The remaining 20% of the dataset is used for testing.

The results of this experiment show that the number of misclassifications is only 3 out of 305 total testing instances, for a testing accuracy of 99.0%. The experiment therefore demonstrates the potential of the GMLG method to produce high accuracy mappings of novel communication paths into representative channel library modules

V. CONCLUSIONS

This paper has introduced a new approach for integrating field measurements into the modeling and simulation of mechatronic systems that are integrated with wireless communication capability. Novel aspects of the approach include feature extraction and feature clustering, which allow for derivation of representative channels and associated channel library modules in an efficient manner from highly compressed, feature-based representations. Moreover, the derived clusters are used to train a classification model, which can be used to classify novel communication paths (not represented in the field measurements) into the most representative channel library modules for simulation. In the experiments presented in the paper, key parameters of the proposed approach, such as the downsampling factor applied to CIRs, and the number of clusters to generate, were derived empirically. A useful direction for future work is the development of automated methods and supporting tools for setting these parameters.

DISCLAIMER

REFERENCES

[1] A. A. Kumar S., K. Ovsthus, and L. M. Kristensen, "An industrial perspective on wireless sensor networks — a survey of requirements, protocols, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1391–1412, 2014.

[2] H. Li, J. Geng, Y. Liu, M. Kashef, R. Candell, and S. Bhattacharyya, "Design space exploration for wireless-integrated factory automation systems," in *Proceedings of the IEEE International Workshop on Factory Communication Systems*, 2019, pp. 1–8.

[3] K. Wiklundh, "Interference challenges for industry communication," 2019, PDF presentation slides from keynote at WFCS 2019, downloaded from https://www.miun.se/en/thank-you-for-participating on 01/22/2020.

[4] J. Geng, H. Li, M. Kashef, Y. Liu, R. Candell, and S. S. Bhattacharyya, "Integrating field measurements into a model-based simulator for industrial communication networks," in *Proceedings of the IEEE World Conference on Factory Communication Systems*, 2020, pp. 1–8.

[5] M. Kashef, R. Candell, and Y. Liu, "Clustering and representation of time-varying industrial wireless channel measurements," in *Proceedings of the Annual Conference of the IEEE Industrial Electronics Society*, 2019, pp. 2823–2829.

[6] R. Candell, M. Kashef, Y. Liu, K. B. Lee, and S. Foufou, "Industrial wireless systems guidelines: Practical considerations and deployment life cycle," *IEEE Industrial Electronics Magazine*, vol. 12, no. 4, pp. 6–17, 2018.

[7] J. Medbo and P. Schramm, "Channel models for HIPERLAN/2 in different indoor scenarios," Ericsson Radio Systems AB, Tech. Rep. 3ERI085B, 1998.

[8] V. Erceg, "TGn channel models," IEEE P802.11 Wireless LANs, Tech. Rep. IEEE 802.11-03/940r4, 2004.

[9] T. Abbas *et al.*, "Simulation and measurement-based vehicle-to-vehicle channel characterization: Accuracy and constraint analysis," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 7, pp. 3208–3218, 2015.

[10] J. Peil *et al.*, "Channel modeling and performance of Zigbee radios in an industrial environment," National Institute of Standards and Technology, Tech. Rep., September 2016.

[11] M. A. Ahmed, W.-H. Yang, and Y.-C. Kim, "Simulation study of communication network for wind power farm," in *Proceedings of the International Conference on Information and Communication Technology Convergence*, 2011.

[12] Y. Liu, R. Candell, K. Lee, and N. Moayeri, "A simulation framework for industrial wireless networks and process control systems," in *Proceedings of the IEEE World Conference on Factory Communication Systems*, 2016, pp. 1–11.

[13] R. Patidar, S. Roy, T. R. Henderson, and A. Chandramohan, "Link-to-system mapping for ns-3 Wi-Fi OFDM error models," in *Proceedings of the Workshop on ns-3*, 2017, pp. 31–38.

[14] M. C. V. Nascimento and A. C. P. L. F. de Carvalho, "Spectral methods for graph clustering — a survey," *European Journal of Operational Research*, vol. 211, no. 2, pp. 221–231, 2011.

[15] P. J.Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[16] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[17] "Networked control systems group — measurement data files," https://www.nist.gov/el/intelligent-systems-division-73500/networked-control-systems-group/measurement-data-files, 2020, visited in January 2020.

[18] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.