Supplemental Material for "LWS: A Framework for Log-based Workload Simulation in Session-based SUT"

1. LWS DSL

We design a DSL to describe the workload features in a human-readable way so that the simulated workload can be executed easily even by a non-expert who is not familiar with the LWS framework. We divide the workload features into two categories, namely the user behavior features and the intensity features. The example of DSL input files have been displayed in Table. 1. As we can see from the table, the user behavior features consist of the index, the name, the API detail including the request method (GET/POST/PUT/DELETE) and the request URL, the parameter including the name and values, and the transition probability between two adjacent nodes in the relational model. Specially, two dummy nodes u_s (INITIAL) and u_e (TERMINAL) are defined. In the intensity features, the mandatory information includes the interval δ to represent the time span in seconds and the exact workload intensity modeling method (reproduction/fitting/generation). According to the workload intensity method, the corresponding components of time series to describe $I(t)$ from different perspectives should be provided. The DSL grammar is defined in the Backus Normal Form (BNF) as shown in Figure 1.

2. Fitting Mathematical Expressions

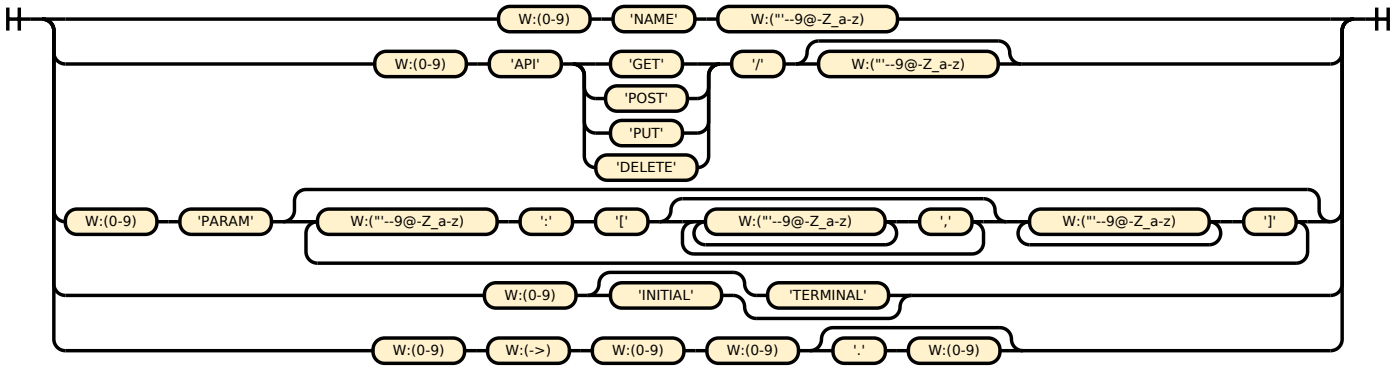
Table 2 shows the detailed mathematical expressions and evaluation metrics including the sum of squares error (denoted by SSE), R-squared value (denoted by R-square), and the root mean square error (denoted by RMSE) of fitting the original intensity. Each mathematical expression is the most suitable one automatically chosen based on RMSE.

3. Generation Parameters

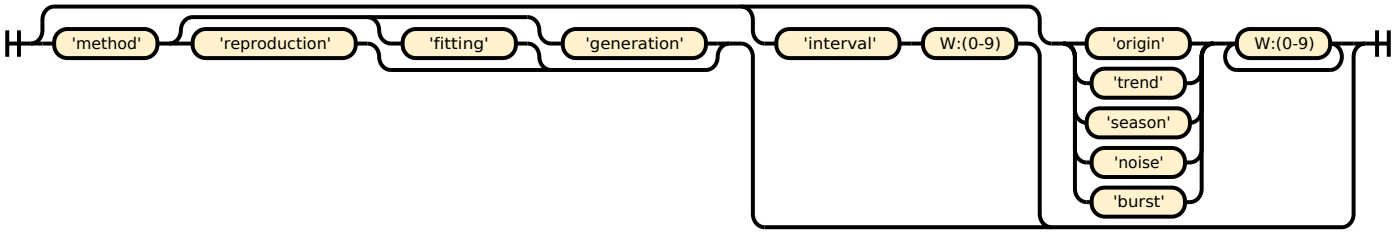
Table 3 shows the parameters and interpolation functions adopted in *Generation(LIMBO)* and *Generation(TSAGen)*.

Table 1: The example of DSL input files (The symbol of ... represents omissions to satisfy the space constraints).

Category	Content
User Behavior Features	0 NAME adding_to_cart 0 API POST /cart 0 PARAM product_id: ['0PUK6V6EV0','6E92ZMY-YFZ','2ZYFJ3GM2N',...] quantity: [1,2,3...] 1 NAME home 1 API GET / 1 PARAM 2 NAME placing_order 2 API POST /cart/checkout 2 PARAM email: ['someone@example.com'] ... ... 6 TERMINAL 7 INITIAL 0->5 1.00 1->3 0.33 1->4 0.67 ...
Intensity Features	method generation interval 10 trend 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 season 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 noise 1 0 2 0 1 0 0 1 1 0 2 0 1 0 1



(a) BNF for user behavior features.



(b) BNF for workload intensity features.

Figure 1: The BNF structures for our DSL grammer.

Table 2: The mathematical expressions of fitting the original intensity in the dataset $\mathcal{B}$.

Method	Component	Mathematical Expression	SSE	R-square	RMSE
Fitting (Origin)	-	$f(x) = a_0 + a_1 \times cos(wx) + b_1 \times sin(wx) + a_2 \times cos(2wx) + b_2 \times sin(2wx) + a_3 \times cos(3wx) + b_3 \times sin(3wx) + a_4 \times cos(4wx) + b_4 \times sin(4wx) + a_5 \times cos(5wx) + b_5 \times sin(5wx) + a_6 \times cos(6wx) + b_6 \times sin(6wx) + a_7 \times cos(7wx) + b_7 \times sin(7wx) + a_8 \times cos(8wx) + b_8 \times sin(8wx), w = 0.04363, a_0 = 34.74, a_1 = -3.123, b_1 = -21.09, a_2 = 7.203, b_2 = -13.5, a_3 = 0.4789, b_3 = -6.229, a_4 = -3.534, b_4 = -3.381, a_5 = -1.036, b_5 = -0.8598, a_6 = -0.4041, b_6 = -0.803, a_7 = 0.1615, b_7 = -0.6252, a_8 = -1.054, b_8 = 0.1594$	2.112×10^4	0.9537	4.319
Fitting (Decomposition)	*trend*	$f(x) = p_1 \times x^8 + p_2 \times x^7 + p_3 \times x^6 + p_4 \times x^5 + p_5 \times x^4 + p_6 \times x^3 + p_7 \times x^2 + p_8 \times x + p_9, p_1 = -3.679 \times 10^{-21}, p_2 = 1.483 \times 10^{-17}, p_3 = -2.332 \times 10^{-14}, p_4 = 1.798 \times 10^{-11}, p_5 = -6.92 \times 10^{-9}, p_6 = 1.19 \times 10^{-6}, p_7 = -7.02 \times 10^{-5}, p_8 = 0.002077, p_9 = 33.61$	79.53	0.9922	0.264
	season	$f(x) = a_0 + a_1 \times cos(wx) + b_1 \times sin(wx) + a_2 \times cos(2wx) + b_2 \times sin(2wx) + a_3 \times cos(3wx) + b_3 \times sin(3wx) + a_4 \times cos(4wx) + b_4 \times sin(4wx) + a_5 \times cos(5wx) + b_5 \times sin(5wx) + a_6 \times cos(6wx) + b_6 \times sin(6wx) + a_7 \times cos(7wx) + b_7 \times sin(7wx) + a_8 \times cos(8wx) + b_8 \times sin(8wx), w = 0.04394, a_0 = 0.001179, a_1 = 0.5394, b_1 = -20.54, a_2 = 11.22, b_2 = -9.706, a_3 = 3.663, b_3 = -4.802, a_4 = -0.2356, b_4 = -4.556, a_5 = 0.134, b_5 = -1.233, a_6 = 0.5156, b_6 = -0.7415, a_7 = 0.6966, b_7 = -0.061, a_8 = -0.2362, b_8 = -0.7853$	296.8	0.9993	0.5121
	noise	$f(x) = a_1 \times exp(-(x - b_1)^2/c_1^2), a_1 = 24.75, b_1 = 1.147 \times 10^3, c_1 = 10.76$	1.403×10^4	0.2838	3.498
	merge	$f(x) = trend + season + noise$	1.405×10^4	0.9692	3.541

Table 3: The parameters of *Generation*.

Method	Component	Parameter	Value
Generation(LIMBO)	*trend*	η_1	1
		η_2	1296
		c_1	27.263
		c_2	40.207
		c_3	26.494
		g_1	linear
	season	η_2	1296
		η_3	9
		η_4	1152
		c_4	-24.975
		c_5	-24.975
		c_6	17.596
		c_7	19.723
		g_2	quadratic
	burst	η_5	104
		η_6	160
		η_7	32
		c_8	32
		g_3	quadratic
	noise	η_8	0
		η_9	6
Generation(TSAGen)	*trend*	θ_1	31.550
		θ_2	0.005
	season	θ_3	65
		θ_4	144
		θ_5	8
		k_1	0
		k_2	0
		d_1	10
		d_2	2
	noise	θ_6	0.206
		θ_7	0.139
		θ_8	4.057