Online Appendix:

A.1 Data Extraction form

Table A.1. Data type and item extracted from each study

Data Type	ID	Data Item	Description
Context	D1-D7	Title, author, venue, publication year, publisher, summary and open challenges	Title, author, venue, publication year, publisher, summary including aim, strength, and weakness of the study and open challenges to be resolved in future.
RQ1	D8-D12	NLP method, feature extraction method, learning type, classifier and HIDS type	NLP method and feature extraction method (automatic or manual) used to implement NLP-based HIDS. Type of learning method, classifier (e.g., recurrent neural network) and type of HIDS (misuse, anomaly) used for intrusion detection in HIDS.
RQ2.1	D13-D14	Attack detection /classification and Attack instances	Attack detection (e.g., benign, malicious) or classification (detect specific attack) and attack instances that are targeted to be detected by NLP-based HIDS
RQ2.2	D15-D16	Dataset and Dataset availability	Dataset used for training or evaluating HIDS, Dataset availability
RQ2.3	D17	Evaluation Metric	Metrics used for evaluating NLP-based HIDS

A.2 Data source, availability, dataset type, and instance mapped with reviewed studies

Table A.2. Dataset type, availability and instance mapped with reviewed studies

Availability	Dataset Type	Instance	Study Ref
public	Real	AWSCTD	S4, S10
		PUS	S58
	Sim	UNM	S7, S16, S18, S23, S27, S30, S49, S52, S53, S56, S58, S60, S61
		Firefox DS	S23
		DARPA	S18, S30
		PLAID	S21
		LID-DS	S9
		NGIDS-DS	S1, S9, S17

	Hyb	ADFA-LD	S1, S2, S3, S5, S6, S7, S8, S11, S14, S15, S16, S17, S18, S19, S20, S21, S22, S23, S24, S25, S28, S29, S30, S31, S35, S37, S38, S40, S41, S42, S43, S44, S45, S46, S47, S48, S50, S51, S54, S55, S57, S59, S62, S63
		ADFA-WD	S1, S6, S12, S42, S54
		ADFA-WD:SAA	S12
		CANALI-WD	S35, S64
private	Real	Customized	S13, S26, S32, S34, S36, S39
	Sim	Customized	S24, S33

A.3 Evaluation Metrics of intrusion detection with mapped studies

Table A.3. Evaluation Metrics of HIDS with mapped studies

Detection Performance	Detection Rate (Recall, detection accuracy, TPR, true positive rate)	S2, S3, S4, S5, S6, S8, S9, S11, S14, S15, S17, S18, S19, S20, S21, S22, S23, S25, S27, S28, S29, S30, S31, S32, S34, S35, S37, S40, S41, S42, S43, S46, S47, S48, S49, S50, S51, S52, S53, S55, S56, S57, S59, S60, S62
	False Alarm Rate (FAR, FPR, false positive rate)	S4, S5, S6, S7, S8, S9, S11, S14, S15, S17, S18, S21, S22, S23, S24, S25, S27, S28, S29, S30, S31, S32, S35, S37, S38, S40, S41, S42, S43, S47, S49, S50, S51, S52, S53, S55, S56, S58, S60, S62
	Receiver Operating Characteristic curve (ROC)/ area under the curve (AUC)	S1, S2, S5, S7, S8, S10, S11, S12, S15, S16, S18, S20, S21, S28, S29, S30, S31, S35, S37, S38, S40, S41, S42, S43, S44, S45, S48, S50, S55, S57, S61, S62, S64
	False Negative Rate (FNR) = Missing Rate	S4, S22, S32, S37, S40, S43, S52, S60
	True Negative Rate (TNR)	S32, S37, S40, S58
	Confusion matrix	S4, S40, S42
	Classification Accuracy or Classification rate (CR)	S3, S4, S8, S10, S11, S12, S18, S20, S22, S36, S37, S42, S43, S46, S48, S54, S55, S57
	Precision	S3, S4, S5, S6, S19, S22, S37, S40, S42, S43, S46, S55, S57
	F-measure	S3, S4, S5, S6, S7, S19, S22, S32, S37, S40, S42, S45, S46, S55, S57
	Classification Error	S4
	Mean error rate	S59
	Matthews Correlation Coefficient (MCC)	S4

Computation Performance	Time (training time/ testing time/ execution time)	S2, S3, S4, S5, S18, S22, S23, S34, S36, S38, S43, S55, S58, S59, S64
Performance of Intermediary Task of Sequence Prediction	BLEU Score	S5, S44, S63
	TF-IDF	S44, S63
	Cosine Similarity	S44, S63