

Table S1. Data description

Variable Name	Variable Description
timestamp	Time the event occurred, measured in seconds since system boot. Helps reconstruct sequences and attack timelines.
processId	Unique ID of the process generating the log. IDs 0–2 are system processes; others are created dynamically.
threadId	ID of the thread within the process that executed the event.
parentProcessId	ID of the parent process that created the current process. Useful for detecting unusual process-spawning behavior.
userId	Numeric ID of the user running the process. System activity uses IDs < 1000; external users or attackers use IDs ≥ 1000.
mountNamespace	Indicates the filesystem namespace the process used. The common value 4026531840 represents standard access; other values can suggest unusual permissions.
processName	Name of the command or program executed (e.g., “bash”, “systemd”). Often important for detecting attacker activity.
hostName	Name of the honeypot machine that generated the log, used to distinguish events across different servers.
eventId	ID of the system event or system call (e.g., fork, exec, clone).
eventName	Human-readable description of the event or system call (e.g., “clone”, “kill”, “execve”). Shows what action was performed.
argsNum	Number of arguments passed to the event. Very large or unusual counts may indicate exploitation behavior.
returnValue	Result of the system call. Usually 0 for success; negative values show errors or failed attempts.
stackAddresses	List of memory addresses relevant to the process at the time of the event. Provides insights into memory usage.
args	List of dictionaries containing details such as file paths, device IDs, process references, or other system-call parameters.
sus	Label indicating suspicious activity (1 = suspicious, 0 = normal).
evil	Label indicating confirmed malicious activity (1 = malicious, 0 = benign)

Table S2. Models parameters and their values

Model	Parameter	Value
KMeans	n_clusters	2
	n_init	10
	random_state	42
	initialization	k-means++ (sklearn default)
	convergence tolerance	1e-4 (sklearn default)
	scaling	StandardScaler
	anomaly score	min distance to nearest centroid (higher = more anomalous)
GMM	n_components	2
	covariance_type	full
	n_init	3
	random_state	42
	regularization (reg_covar)	1e-6 (sklearn default)
	convergence tolerance	1e-3 (sklearn default)
	anomaly score	negative log likelihood (higher = more anomalous)
Spectral Clustering	n_clusters	2
	affinity	nearest_neighbors
	n_neighbors	15
	random_state	42
	training subsample	30,000 events (RandomState(42))
	scaling	StandardScaler
	anomaly score	distance to nearest training centroid (higher = more anomalous)
DBSCAN	eps	0.8
	min_samples	10
	metric	Euclidean (sklearn default)
	algorithm	auto (sklearn default)
	n_jobs	-1
	training subsample	30,000 events (RandomState(42))
	scaling	StandardScaler
	anomaly score	distance to nearest core point via NearestNeighbors (higher = more anomalous)
Hierarchical Clustering	n_clusters	2
	linkage	Ward
	random_state	N/A (AgglomerativeClustering has no random_state; subsampling used RandomState(42))
	training subsample	30,000 events (RandomState(42))

	scaling	StandardScaler
	anomaly score	distance to nearest training centroid (higher = more anomalous)
Isolation Forest	n_estimators	100
	contamination	tuned on validation set; optimal value = 0.001
	random_state	42
	n_jobs	-1
	scaling	StandardScaler
	anomaly score	negated decision function (higher = more anomalous)
VAE-DoSE	input_dim	8
	latent_dim	8
	encoder	Linear(8→64) → ReLU → Linear(64→mu/logvar)
	decoder	Linear(8→64) → ReLU → Linear(64→8)
	optimizer	Adam
	learning_rate	1e-3
	batch_size	256
	epochs	25
	loss function	MSE reconstruction + KL divergence
	early stopping	none
	DoSE density estimator	Gaussian KDE, bandwidth = $0.2 \times \text{std}(\text{training energies})$
	KDE training samples	up to 100,000 (random subsample of training set)
	DoSE anomaly score	negative log density of energy (recon + KL)
	note	DoSE-SVM variant from Highnam et al. ⁸ was not reproduced; KDE-based density estimation used instead

Table S3. Validation and test evaluation metrics to identify the suspicious (SUS) from normal cyberthreat

Model	Set	PR AUC	ROC AUC
Isolation Forest	Validation	0.0450	0.945
	Test	0.970	0.779

VAE-DoSE	Validation	0.648	0.996
	Test	0.996	0.963
KMeans	Validation	0.988	0.997
	Test	0.995	0.958
GMM	Validation	0.763	0.999
	Test	0.996	0.959
Spectral Clustering	Validation	0.989	0.999
	Test	0.996	0.958
DBSCAN	Validation	0.634	0.930
	Test	0.996	0.971
Hierarchical	Validation	0.989	0.995
	Test	0.996	0.958

Table S4. McNemar's test results between model pairs to predict test SUS attacks

Model A	Model B	p - value	Model A is correct Model B is wrong	Model A is wrong Model B is correct
KMeans	GMM	0.0000	5,801	164,183
KMeans	Spectral	0.0000	11,874	171,964
KMeans	Hierarchical	0.0001	7,183	6,710
KMeans	Isolation F.	0.0000	1,853	2,824
KMeans	VAE-DoSE	0.0000	7,217	91,627
GMM	Spectral	0.0000	9,836	11,544
GMM	Hierarchical	0.0000	158,890	35
GMM	Isolation F.	0.0000	161,359	3,948
GMM	VAE-DoSE	0.0000	74,165	193
Spectral	Hierarchical	0.0000	169,434	8,871
Spectral	Isolation F.	0.0000	169,223	10,104
Spectral	VAE-DoSE	0.0000	84,618	8,938
Hierarchical	Isolation F.	0.0000	3,886	5,330
Hierarchical	VAE-DoSE	0.0000	163	85,046
Isolation F.	VAE-DoSE	0.0000	5,484	88,923

Table S5. Test evaluation metrics to identify the malicious from benign cyberthreat

Model	Set	PR AUC	ROC AUC
Isolation Forest	Test	0.967	0.838
VAE-DoSE	Validation	-	-
	Test	0.971	0.936
KMeans	Validation	-	-
	Test	-	-

	Test	0.999	0.994
GMM	Validation	-	-
	Test	0.999	0.995
Spectral Clustering	Validation	-	-
	Test	0.971	0.874
DBSCAN	Validation	-	-
	Test	0.998	0.992
Hierarchical	Validation	-	-
	Test	0.999	0.994