TrendAI™ Research

Forecasting Future Outbreaks

A Behavioral and Predictive Approach to Proactive Cyber Risk Management

Marco Balduzzi, Roel Reyes, Jessica Balaquit, Ryan Flores, Benjamin Zigh

Contents

Published by
TrendAI™ Research

Written by
Marco Balduzzi
Roel Reyes
Jessica Balaquit
Ryan Flores
Benjamin Zigh

Organizations worldwide face a continuous battle against increasingly sophisticated malware and ransomware attacks, driving up costs and operational downtime. While security platforms are essential, they typically rely on a reactive approach: They wait for a signature of a malicious file to appear before raising alerts or remediating. This strategy ensures incident response but is limited from anticipating threats.

To overcome this limitation, TrendAI™ Research introduces a new approach that accurately forecasts the likelihood of a future malware outbreak on any endpoint up to 30 days in advance. This capability is built on the critical principle that many threats do not occur randomly but are directly influenced by the behavior of their victims. TrendAI™ Research analyzed high-volume, real-time activity data, including web traffic, application usage, and user interactions to statistically model and quantify risk.

This innovation lies in two distinct, yet integrated, analysis engines. The first engine uses established statistical methodologies to link specific user actions (such as visiting unclassified websites or rapidly installing new software) to future infections for different threat categories. This first engine provides explainable reasoning. The second engine refines the insights using advanced modeling, enabling TrendAI™ Research to simultaneously predict the probability of multiple, distinct threats (such as ransomware, trojans, or potentially unwanted applications or PUAs) hitting a specific machine.

This strategic risk assessment allows security leaders to move beyond general defense. Instead of treating all machines equally, they can prioritize resources where the risk is highest and most imminent. For instance, if the model forecasts a heightened risk of ransomware on critical departmental servers, immediate, targeted mitigation steps like stronger patch enforcement or network segmentation can be taken before the outbreak even begins. By shifting focus from simply cleaning up incidents to preemptive risk reduction, this solution changes how defenses are deployed and better operationalizes security investment.

By combining predictive risk modeling with automated protection, organizations are enabled to quantify, manage, and reduce cyber risk. The methodologies described in this research are actively driving the evolution of TrendAI Vision One™ Cyber Risk Exposure Management (CREM),[1] providing a foundation for proactive security strategies that address the risk of future malware outbreaks.

Introduction

Malware outbreaks continue to represent a major problem for organizations worldwide[2, 3] with ransomware infections alone costing affected organizations millions of dollars on average.[4] Various forms of malware continue to be leveraged by criminals to enable their attacks: Trojans, for example, are used to maintain persistence into compromised networks, while hacking tools or "hacktools" are used to conduct lateral movement from initial infection to get to the cybercriminals' targets, such as confidential information stored in internal databases.

To protect organizations against cybercrime, the industry has introduced several security solutions over the years, with anti-malware playing a central role in defending against outbreaks. This involves detecting the presence of malicious software on an endpoint, such as upon download, and then quarantining it. Intrusion detection system (IDS) solutions and network detection response (NDR) platforms inspect network traffic through known signatures of attack patterns to detect anomalies, which could be due to the presence of an infection or ongoing attack. More recently, in response to the fact that modern attacks are becoming more sophisticated and are challenging the capabilities of traditional solutions,[5] platforms such as endpoint detection response (EDR) have been progressively introduced in the market.

These solutions, when combined, provide a comprehensive strategy against cyberattacks that employ a reactive approach. Anti-malware, anti-spam, web application firewalls, and IDS solutions all operate by waiting for evidence of an attack before triggering an alert or initiating remediation. While this is effective, reliable, and essential in most cases, it does not allow an organization to strategically plan its defense. In contrast, a proactive approach enables an organization to anticipate potential outbreaks and stay ahead of future attacks performed by cybercriminals.

In this work, researchers from TrendAI™ Research build upon their previous study[6] and introduce a model that estimates the risk of future malware outbreaks. The new approach not only forecasts potential outbreaks within the next 30 days but also provides a breakdown of the risk probabilities by malware class. Additionally, it offers explainable reasoning behind these predictions.

Being able to profile the risk for different classes of malware enables organizations to better strategize their defenses, and this is very important because each organization is susceptible to different classes of attacks. For example, hospitals are known to be a primary target for ransomware due to the sensitive nature of the data they handle. Similarly, each machine is also susceptible to different classes of infection, like how Windows desktops operated by regular employees in an organization are often targeted by PUAs or adware.

The new approach that TrendAI™ Research introduces consists of employing behavioral user data to estimate the risk of future attacks. Studies[7] have previously shown that improper user behaviors, such as visiting shady websites, put users at risk of infection. This new approach builds on top of these findings by extending the methodology and introducing a system for analyzing real-time data at scale.

This new approach forecasts future malware outbreaks on endpoints up to 30 days in advance. This shifts security strategy from reactive incident response to proactive threat anticipation.

The approach provides risk scores by specific malware class (ransomware, trojan, PUA) and offers explainable reasoning linking the risk directly to specific user behaviors. This allows organizations to strategically prioritize, and tailor defenses based on the type of threat they are most vulnerable to.

This large-scale study, which involves over 10.7 million endpoints, demonstrates the system's effectiveness and readiness for deployment in complex, global enterprise environments such as CREM.

Overview

Given a time t, this approach consists of training a system on the past activities of the users (at t – *30 days*) and using current and future malware incidents (at t + *30 days*) for labeling. This strategy enables the system to estimate a prediction for an endpoint e at time t for the upcoming window.

The system operates in two phases. As shown in Figure 1, one is used for training and one for estimating the risk of future malware outbreaks.

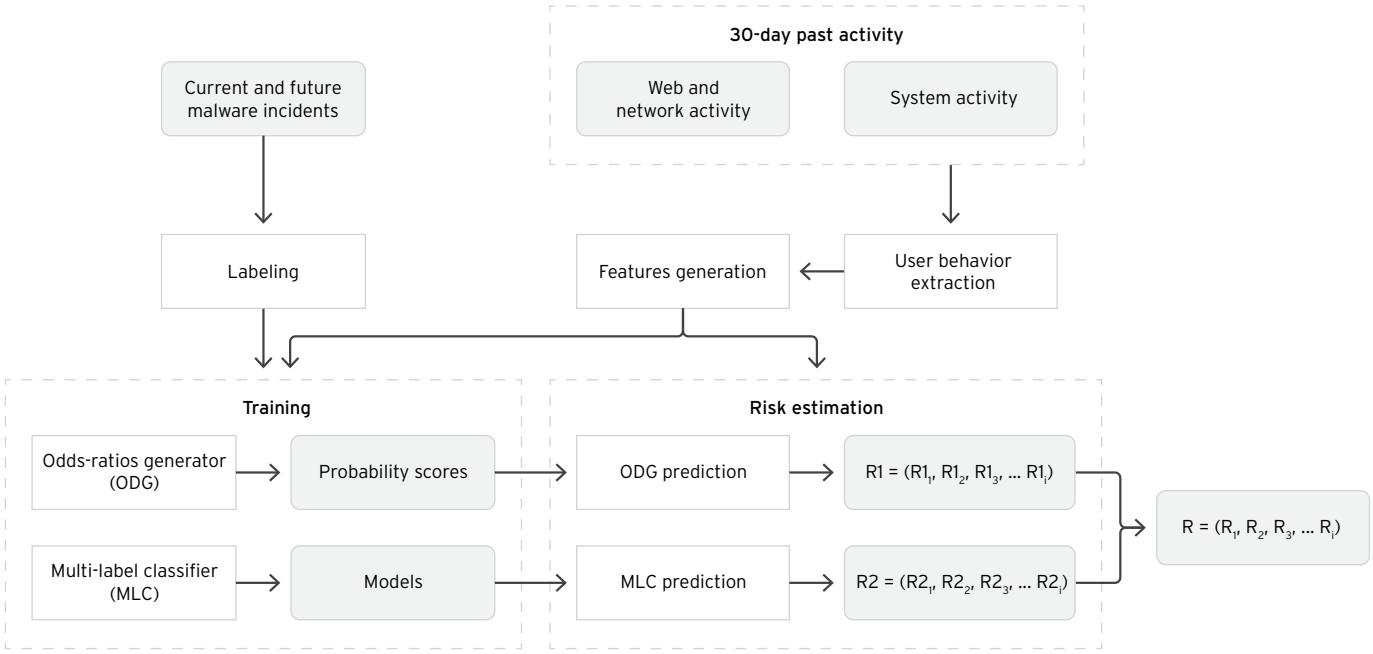


Figure 1. System architecture of the new approach proposed by this study

The two macro components are the odds-ratios generator (abbreviated as ODG, rather than ORG, for the purposes of this research) and the multi-label classifier (MLC).

The ODG is built on top of a previous study[8] that extends with the risk estimation module (ODG prediction in Figure 1) and other internals that are detailed later in the subsection "Design: Odds-Ratios Generator (ODG)." This component is combined with the MLC, which leverages supervised-learned models to refine the risk estimation R. This is made for each malware class under analysis, which in this current setup consists of "coinminer," hacktool, PUA, ransomware, trojan, and virus as per the definitions in Table 1.

Malware class	Description
Coinminer	Mines cryptocurrencies
Hacktool	Used for local hacking or lateral movement
PUA	Potentially unwanted application
Ransomware	Prevents user access to files allowing threat actors to demand ransom for its return
Trojan	Disguises as a legitimate program
Virus	Infects endpoints automatically

Table 1. Malware classes considered in this work and their definitions

Both these macro-components (the ODG and the MLC) rely on a preprocessing phase in which activity data is loaded from TrendAI™ telemetry and preprocessed accordingly. The first component extracts the users' behavior from the activity of the users on the endpoints, which includes, but is not limited to, the following data:

- Network traffic generated from the applications installed on the machines
- Software regularly downloaded from the internet
- Categories of software executed and websites visited
- The frequency of machine utilization

This information is passed to the second component, which generates the features used by the ODG and the MLC. Note that both these components rely on the same set of features (presented in Table 4), but the way in which these are used is different, which will be discussed later.

At the same time, another component labels the endpoint as "infected" or "not infected" (and specifies the related malware class) based on whether the endpoint encounters malware in the following 30 days. Since the goal is to capture the effect where the user's behavior triggers a malware infection (behaviors such as visiting shady websites or installing untrusted applications), this approach focuses exclusively on behaviors directly related to the user, disregarding any automated executions by applications. In addition, this approach considers only first-stage infections, that is, it disregards endpoints that were previously infected. The labeling is used in the training phase by the ODG for generating the case-control groups and by the MLC to train the models.

Meanwhile, the features are used both in the training phase as well as during the risk estimation phase to compute the risk score for a given endpoint e at real time. In this phase, the system operates by fetching the activity of the endpoint's user from the telemetry in real time, extracting the user's behavior, generating the features, and using that information as input for the system to generate the risk score $R = \{R_1, R_2, \ldots R_i\}$, where i represents the number of malware classes taken into consideration as an average of the risk scores estimated by the individual components (the ODG and the MLC). In this work, there are six malware classes considered, which will be enumerated in the "Training" section.

Design: Odds-Ratios Generator (ODG)

The odds-ratios generator (ODG) component is tasked with generating the odds ratios for the dataset under analysis, which, in the case of this approach, consists of behavioral user data and malware infection.

In statistics, odds ratios are mathematical formulations that quantify the association between two events. For better understanding, let's make an analogy within the medical domain: the likelihood that smoking a high quantity of tobacco per day would result in lung cancer. The process of generating an odds ratio consists of dividing the training population into two groups: One group is made up of individuals exposed to risk, and the other is made up of those who are not. These groups are then divided into two more groups: those who later developed cancer, and those that did not. This process results in a so-called contingency table, which is used as input to compute the odds ratio.

Table 2 provides an example that applies to this work: The case group is composed of endpoints that encountered ransomware $(a + c)$, while the control group consists of endpoints that did not encounter the malware $(b + d)$. The odds ratio OR is the ratio of the odds of exposure among cases $\frac{a}{c}$ to the odds of exposure among controls $\frac{b}{d}$. As an example, an OR of 2.3 indicates that visiting more than 200 domains per day increases the probability of being infected by ransomware by 230%.

Visited domains per day (average)	Encountered ransomware	
	Yes	No
> 200	a	b
≤ 200	c	d

Table 2. Example of a contingency table used as input for generating the odds ratios

During training, this process is computed through all the features and all malware classes. The implementation of this execution is described in the "Training" section, and the results are presented in Tables 4 and 5.

Following the training phase, the same component is used to estimate the risk of endpoints at real time. In this configuration, the system operates by fetching the past 30 days of endpoint activity from the telemetry, extracting the user behavior, and computing the features. The estimate risk R_i for a class of malware i is computed with the following formula:

$$S_i = \sum \left(\log(\text{OR}_j) \times \text{SignificantFeature}_j \right)$$

$$R_i = \frac{1}{1 + e^{-S_i}}$$

Note that only the features that are statistically significant are used (i.e., those with p-value < 0.05 in this implementation, which will be explained later).

Design: Multi-Label Classifier (MLC)

The multi-label classifier (MLC) component implements a collection of independent binary classifiers, each trained to distinguish between a specific malware class behavior and normal (control) endpoint behavior. Unlike traditional multiclass classification approaches where the probabilities across all classes sum to 1, this architecture treats each malware prediction task as an independent binary classification problem. This design enables the simultaneous prediction of multiple malware classes and delivers estimations that can collectively exceed unity when multiple threats are simultaneously present.

The training process begins by constructing datasets for each malware class using a 1:2 ratio of endpoints that encountered malware to normal samples. For a given malware class i with n_i endpoint samples, a corresponding dataset was created by pairing all n_i malware encountering samples with a randomly sampled subset of $2n_i$ control samples from the control population. This approach ensures class balance while preserving the statistical properties of normal endpoint behavior. The sampling is performed without replacement across different malware classes to maintain dataset independence and prevent overfitting to specific normal behavior patterns. For computational efficiency, control endpoints are randomly sampled, acknowledging that this might introduce some demographic bias compared to the approach adopted by the ODG (case-control groups).

During the risk estimation phase, the system processes endpoint behavioral data through all trained classifiers simultaneously. Following the random forest methodology proposed by Breiman,[9] each independent binary classifier for malware class i outputs a risk score R_i representing the likelihood that the endpoint exhibits behavior consistent with that specific malware encountered. The risk score is computed as the proportion of trees in the forest that vote for the positive class (malware), which converges to the true probability as the number of trees approaches infinity. Each risk assessment is computed independently as:

$$R_i = \frac{1}{T}\sum_{t=1}^{T} I\left(h_t(x) = i\right)$$

where R_i is the independent risk score for malware class i, T is the number of trees in the forest, $h_t(x)$ is the prediction of the t-th tree for input x, and $I(\bullet)$ is the indicator function.

This approach allows for simultaneous detection of multiple malware classes, as each classifier operates independently without requiring probabilities to sum to unity, enabling security analysts to assess and prioritize response efforts based on individual threat-specific risk scores.

Datasets Overview

This section gives an overview of the datasets used both for computing the odds ratios and for training (and testing) the MLC. The datasets are collected by TrendAI™ telemetry though a series of sensors that are integrated in anti-malware and similar endpoint solutions deployed on Windows endpoints.

The process is the same for all endpoints, and the data format is consistent across the different versions of the solutions. Overall, the following telemetry datasets were leveraged:

- The first dataset (named *WRS*) consists of information associated with the web reputation system of security solutions. This dataset includes the URLs contacted by an endpoint, such as those using a browser or any HTTP-enabled application, together with the timestamp and the associated user-agent. Each request is associated with both a safety score and a category score assigned by an internal classification mechanism known as web reputation system (WRS).

  The safety score indicates the risk level associated with the visited website and takes the following values: "normal" (non-malicious website), "suspicious" (potentially linked to spam or compromise), "dangerous" (confirmed to be malicious), or "unclassified".

  The category score indicates the category the contacted URL belongs to. For example, the category "Adult" is associated with sites that might be considered inappropriate for children, while "Business" includes sites related to business, employment, or commerce.

  To determine the popularity of the websites visited, this research relies on Tranco Top 10k.[10]

  Note that since this study is focused on collecting activities from the users only, this dataset is processed by the system to retain only user-generated traffic, but this excludes automated-generated traffic like operating system (OS) updates. (This is discussed further in the "Training" section.)

- The second dataset (named *CS*) includes information on all binaries (i.e., .exes) executed on the machine and the associated loaded dynamic libraries (i.e., dlls) together with their timestamp, signer information, and version.

  This research also collected information on the OS, the country, and the industry (e.g., healthcare or manufacturing) of the organization in which the endpoint is deployed.

  This dataset was enriched with information about the application category tied to an executable. For this, the approach of Bilge et al.[11] was adopted and the executable names were compared against Capterra, a service that classifies popular enterprise software, to identify their categories.

- The third dataset (named *VS*) consists of malware events extracted from security incidents detected on the endpoint by the security solutions. Some examples are malware dropped through drive-by downloads or attached to emails. This information includes the name of the detected malware, including class (e.g., ransomware), SHA-1, and timestamp. CARO's naming convention[12] was adopted for this.

Training

This section discusses how the training for this new approach was conducted. For both the ODG and the MLC, an initial set worth two months of data (November and December 2024) was employed, totaling a number of 12,320,026 endpoints. These endpoints consisted of Windows versions 7, 8, 10, 11, and Vista in all their flavors (Home, Pro, Enterprise, etc.) and languages. The dataset covered over 208 countries, making it representative of the general behavior of users worldwide. Figure 2 provides an overview of the top 20 countries or regions by number of endpoints tested.

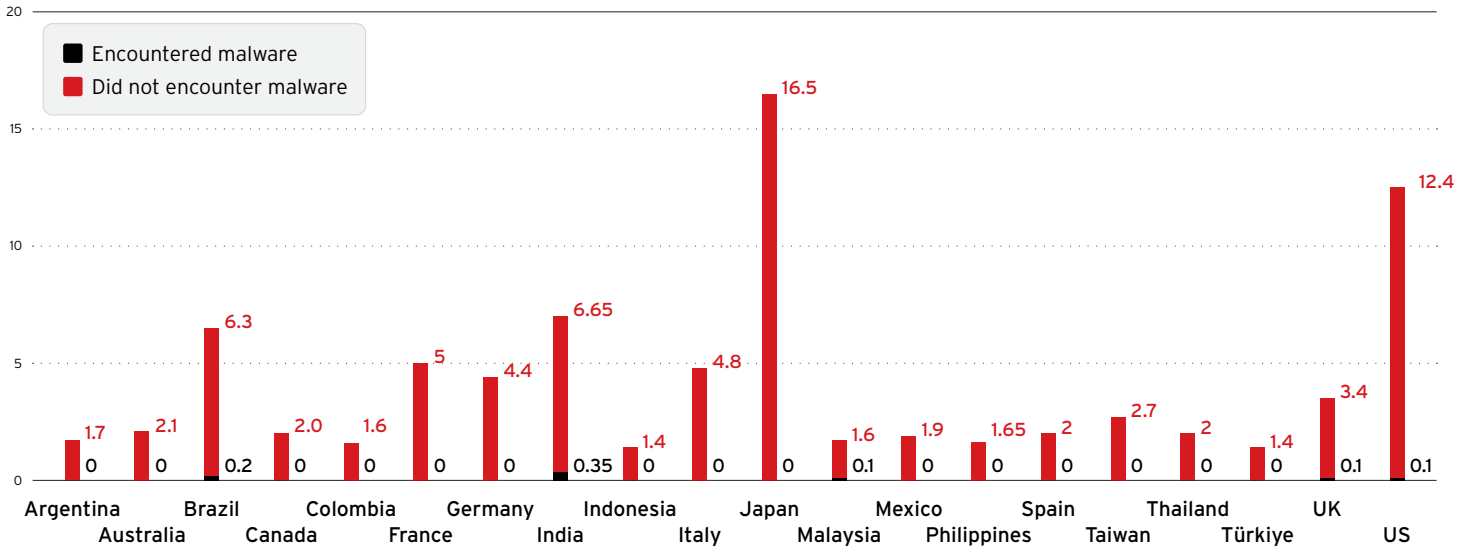


Figure 2. Percentage distribution of endpoints that encountered malware and those that did not across the top 20 countries or regions by number of endpoints tested

For each endpoint, the training set consisted of the user's behavior over the previous 30 days, such as the webpages visited by the user and the applications installed, followed by a subsequent event suggesting a malware incident, such as a binary quarantined by the anti-malware solution. This study did its best to ensure that the events were related to first infections (i.e., excluding malware dropped by previous infections) and to retain only behaviors directly attributable to the user (i.e., excluding any activity performed by software installed on the machine).

Following this preprocessing phase, the cleaned datasets were passed to the ODG and the MLC. The result of the training (in the form of probability scores and models) was used in the risk estimation phase to compute the risk score for a given endpoint e at real time. This is further discussed in the section "Forecasting Future Outbreaks at Scale," where this approach was used to conduct a large-scale study across 10.7 million endpoints in an automated way.

Experimental Setup

In developing this new approach, TrendAI™ Research designed an architecture that guarantees quick training and classification operations. This section outlines the setup, resources, and configuration utilized in this process.

The data repository for the models was stored in Amazon S3 and utilized an EC2 instance for computation. This instance featured 32 cores, 256 GB of RAM, and a primary storage of 1 TB. An additional 2 TB of drive space was allocated for PySpark, which eliminated storage capacity challenges while transferring data from S3 to local storage. The EC2 instance was provisioned with the permissions necessary to access the S3 bucket, essential for PySpark operations. JupyterHub was installed to enable coding and testing, employing Python version 3.12 along with several modules such as *pandas*, *numpy*, *pyspark*, and *s3fs*. This configuration guaranteed efficient data retrieval and processing within the Spark environment.

PySpark was also optimized at different stages of data processing, including collection, preparation, sanitation, feature extraction, and training, with the goal of reducing memory allocation and improving file handling capabilities. For example, both the executor memory and executor cores capacity were increased to support operations related to data preprocessing and feature extraction (i.e., +50% and +30%, respectively).

Performance

The collection and preprocessing of the training set took 30 minutes for one day's worth of *WRS* logs, 20 minutes for *CS* logs, and 10 minutes for *VS* logs, totaling 61 hours (approximately 2.5 days) for the entire training set. Extracting user behaviors and generating features required an additional 1.5 days. The ODG component took two days to generate the case-control groups and an additional day to compute the odds ratios.

The MLC component exhibited the following performance metrics. Hyperparameter tuning, which optimized parameters for each malware class independently using cross-validation, required a total of 10 hours across all six malware classes. The data processing and training phases for each independent classifier took approximately seven minutes. This duration included feature preparation, dataset balancing, and the training process itself. The per-classifier training time supports a model-rolling architecture, allowing classifiers to be retrained as new training data becomes available.

During prediction, the system estimates risk scores R_i with an average time of 155 ms per endpoint e and malware class i. This processing time results from optimizations including vectorization of the feature scaling step. Thus, it can be concluded that the prediction time supports real-time risk estimation deployment.

Implementation: Odds-Ratios Generator (ODG)

As discussed in the overview, the adopted approach for generating the odds ratios – i.e., the *training* phase of the odds-ratios generator (ODG) component – followed previous work. Some notable improvements included extending the labeling period D_p from two weeks to one month (i.e., the month of December), extending the number of features to 28 to better capture the behaviors of the users (including a wider set of 218 countries in the training), and leveraging a wider population of 12,320,026 endpoints and 1,472 organizations.

Following the preprocessing phase, which included only endpoints initially compromised (i.e., endpoints that could have been "re-compromised" due to a previous infection were not considered) and sampled the initial set with a ratio of 1/1,000 to reduce computation overhead, the case-control groups were constructed by considering a ratio of control to cases of three to one, that is, for each case enrolled, three control endpoints matching the following endpoint characteristics: country, type of organization (large or SMB), operating system (flavor, version, and language), and industry sector.

Table 3 provides an overview of the 102,424 case-control groups used for the generation of the odds ratios. While PUAs and trojans accounted for a good number of the groups because of their large presence in the malware ecosystem (likely due to numerous families and high popularity of adoption by attackers), minor and more specialized malware classes such as coinminers and ransomware accounted for about a thousand groups each, making the training set large enough for generating the odds ratios.

	Case	Control	Total
Coinminer	249	747	996
Hacktool	5,148	15,444	20,592
PUA	12,483	37,449	49,932
Ransomware	217	651	868
Trojan	5,781	17,343	23,124
Virus	1,728	5,184	6,912
Total	**25,606**	**76,818**	**102,424**

Table 3. Case-control groups created during training (alphabetically ordered by malware class)

The thresholds used for discriminating between the population exposed to risky behaviors (e.g., visiting a large number of distinct categories of websites per day) and safe behaviors (e.g., visiting trusted websites only) are shown in Table 4. Different features have different thresholds; the method introduced in this study employs both "static thresholds" for features where, for example, visiting even a single malicious website would put the user at risk, and "dynamic thresholds," which are computed by the system using the third quartile of the feature's distribution.

Note that for some features, both types of thresholds are employed. This is because a first pass is conducted to retain only those endpoints that exceed normal activity (e.g., visiting a number of URLs per day that surpasses what would be typical OS updates) and then a second pass is performed to actually identify the endpoints at risk.

As an example, feature #21 captures the number of distinct software applications or products the user used per day. What happens in practice is that if a user regularly employs more than 119 applications, this behavior is considered "exposed" to risk because it is above the threshold. This information is used by the algorithm in generating the contingency table needed for computing the odds ratios (see Table 2).

Feature ID	Description	Category	Static threshold	Dynamic threshold
1	Suspicious and dangerous sites visited per day	Content	> 0	-
2	Presence of shady traffic	Content	> 0	-
3	Undefined or unknown sites visited per day	Content	> 0	> 12
4	Suspicious sites visited per day	Content	> 0	-
5	Peer-to-peer sites visited per day	Content	> 0	> 5
6	Malicious sites visited per day	Content	> 0	-
7	Illegal gambling sites visited per day	Content	> 0	-
8	Illegal sites visited per day	Content	> 0	-
9	Entertainment sites visited per day	Content	> 0	> 9
10	Business sites visited per day	Content	> 0	> 45
11	Explicit adult sites visited per day	Content	> 0	-
12	Gambling sites visited per day	Content	> 0	-
13	Distinct countries the machine is operated in	Diversity	-	> 1
14	Average distinct SHA-1 by path (indicative of applications installed and run)	Diversity	-	> 7.76
15	Average distinct domains visited	Diversity	> 4.5	> 51.45
16	Distinct domains visited per day	Diversity	> 18	> 145
17	Distinct categories of sites visited per day	Diversity	> 2	> 38
18	Ratio not in Tranco diff (monitors if accessed domains are part of the Tranco list)	Popularity	-	> 0.20
19	Product version percentile rank	Popularity	-	> 0.75
20	Ratio not in Tranco (see feature 18)	Popularity	> 1%	> 15.24%
21	Distinct software applications or products used per day	Volume	-	> 119.25
22	Installed software applications or products	Volume	-	> 159

Feature ID	Description	Category	Static threshold	Dynamic threshold
23	File signers (indicative of the number of providers of various applications on the machine)	Volume	-	> 31
24	Percentage of endpoint use at nighttime	Volume	-	> 0.85
25	Median file prevalence (indicative of how common applications on the machine are)	Volume	-	> 1.93
26	Percentage of endpoint use at nighttime per day	Volume	> 1%	> 28.51%
27	Refined file paths (indicative of the heterogeneity of applications on the machine)	Volume	-	> 52
28	Transformed median file prevalence (see feature 25)	Volume	-	> 0.07

Table 4. Features employed by the ODG together with the thresholds used to discriminate between behaviors classified as "exposed" and "not exposed"

The last step of the training consisted of computing the odds ratios for the different classes of malware. These probabilities express the "likelihood" that a certain behavior puts the user at risk. The algorithm works by taking the contingency tables for each malware class and risk factors and compute the odds ratio using the χ^2-test. To avoid division by zero in computing the ratios, this study applied Haldane's correction (add 0.5 to all cells if any of them is zero). The study considered the exposure effect to be statistically significant if the p-value obtained from the chi-square test is < 0.05. In addition to the p-value, the 95% confidence interval (CI)[13] was also computed; this provides information about the range in which the odds are within 95% probability. Given a CI, it can be determined if the test is statistically significant if it does not include the value 1, called the "value of zero effect."

The results of the odds-ratio generation are shown in Table 5. For example, using a large variety of software applications (more than 159) increases the probability of encountering backdoored software, such as trojans, by 61% (feature #22). This is understandable because in a desktop environment where users are allowed to download and install any application found over the internet (if no enterprise policy defining application whitelists is in place), there is a higher chance that some of these applications would contain malware.

Similarly, an endpoint found to have installed a multitude of different concurrent signed applications on disk (feature #23), including self-signed ones (i.e., more than 31), has double the chances of encountering hacktools. This is because an attacker could use that machine to disguise their presence and conduct lateral movement.

Interestingly, visiting gambling websites (feature #12) exposes the user to the risk of encountering PUAs, trojans, and hacktools, but not coinminers, ransomware, or virus. This is likely because certain classes of malware are more likely to be associated with certain categories of websites (because of the distribution model of the cybercriminals behind the malware campaigns).

ID	Description	Coinminer	Hacktool	PUA	Ransomware	Trojan	Virus
1	Suspicious and dangerous sites visited per day	4.74 (**)	2.81 (**)	2.98 (**)	2.73	3.97 (**)	2.80 (**)
2	Presence of shady traffic	5.28 (**)	2.81 (**)	2.98 (**)	3.07 (*)	3.97 (**)	2.83 (**)
3	Undefined or unknown sites visited per day	0.71	1.74 (**)	2.15 (**)	1.71 (**)	1.80 (**)	1.13 (*)
4	Suspicious sites visited per day	0.94	1.37 (**)	1.91 (**)	0.68	1.78 (**)	0.77
5	Peer-to-peer sites visited per day	0.56 (**)	1.56 (**)	1.81 (**)	0.85	1.53 (**)	0.80 (**)
6	Malicious sites visited per day	2.21 (**)	2.80 (**)	2.90 (**)	1.48	3.01 (**)	2.52 (**)
7	Illegal gambling sites visited per day	0.91	1.47 (**)	2.00 (**)	0.68	1.83 (**)	0.92
8	Illegal sites visited per day	1.00	8.51 (**)	8.06 (**)	2.99	5.83 (**)	4.28 (**)
9	Entertainment sites visited per day	0.77	1.62 (**)	1.95 (**)	1.22	1.76 (**)	1.05
10	Business sites visited per day	0.58 (**)	1.41 (**)	1.82 (**)	1.45 (*)	1.46 (**)	0.75 (**)
11	Explicit adult sites visited per day	2.31	1.78 (**)	2.37 (**)	1.00	2.29 (**)	1.15
12	Gambling sites visited per day	0.94	1.37 (**)	1.91 (**)	0.68	1.78 (**)	0.77
13	Distinct countries the machine is operated in	0.87	1.03	0.99	0.94	1.18 (*)	1.49 (**)
14	Average distinct SHA-1 by path (indicative of applications installed and run)	1.13	1.26 (**)	1.46 (**)	1.30	1.22 (**)	1.29 (**)
15	Average distinct domains visited	0.77	1.49 (**)	1.77 (**)	1.10	1.53 (**)	0.86 (*)
16	Distinct domains visited per day	0.84	1.50 (**)	1.89 (**)	1.31	1.57 (**)	0.84 (**)
17	Distinct categories of sites visited per day	0.79	1.58 (**)	1.97 (**)	1.47 (*)	1.68 (**)	0.85 (*)
18	Ratio not in Tranco diff (monitors if accessed domains are part of the Tranco list)	1.06	1.08 (*)	1.19 (**)	0.77	1.19 (**)	1.08
19	Product version percentile rank	2.84 (**)	1.01	1.08 (**)	1.71 (**)	1.29 (**)	1.76 (**)

ID	Description	Coinminer	Hacktool	PUA	Ransomware	Trojan	Virus
20	Ratio not in Tranco (see feature 18)	1.17	0.99	1.13 (**)	0.75	1.13 (**)	1.11
21	Distinct software applications or products used per day	1.46 (*)	1.63 (**)	1.78 (**)	2.29 (**)	1.46 (**)	1.80 (**)
22	Installed software applications or products	1.21	1.59 (**)	1.85 (**)	2.45 (**)	1.61 (**)	1.64 (**)
23	File signers (indicative of the number of providers of various applications on the machine)	1.07	2.02 (**)	2.10 (**)	1.37	1.72 (**)	1.83 (**)
24	Percentage of endpoint use at nighttime	1.07	1.13 (**)	1.26 (**)	1.37	1.15 (**)	1.33 (**)
25	Median file prevalence (indicative of how common applications on the machine are)	0.90	0.75 (**)	0.66 (**)	0.47 (**)	0.85 (**)	0.83 (**)
26	Percentage of endpoint use at nighttime per day	1.03	1.02	1.27 (**)	1.92 (**)	1.19 (**)	1.16 (*)
27	Refined file paths (indicative of the heterogeneity of applications on the machine)	1.50 (*)	1.94 (**)	1.95 (**)	2.08 (**)	1.88 (**)	1.84 (**)
28	Transformed median file prevalence (see feature 25)	0.95	1.64 (**)	1.70 (**)	1.73 (**)	1.38 (**)	1.71 (**)

Table 5. Odds ratios generated by the ODG during training. Only the probabilities with p-value less than 0.05 are considered statistically significant and used for generating the risk scores Ri. (*: $p < 0.05$; **: $p < 0.01$)

Implementation: Multi-Label Classifier (MLC)

As discussed in the overview, the multi-label classifier (MLC) component consisted of independent binary classifiers, one for each class of malware. The training methodology followed a similar data preprocessing pipeline of the ODG component, utilizing the same set of 28 behavioral features extracted from endpoint telemetry data.

For training and validation, this study performed an 80/20 split on the November–December 2024 dataset. For testing, the January 2025 dataset was employed after removing the malware-positive samples.

The training process began with hyperparameter tuning performed independently for each malware class, which evaluated different combinations of model parameters using cross-validation. Each binary classifier was then trained independently using the optimized hyperparameters, allowing specialized decision boundaries tailored to each malware class.

Following initial training, comprehensive performance evaluation was performed for each binary classifier through ROC (receiver operating characteristic) curve analysis. This analysis yielded curves representing the true positive rate against the false positive rate across all possible classification thresholds, providing insight into each classifier's discriminative ability. AUC (area under the curve) was calculated for each classifier, serving as a threshold-independent measure of classification performance. As shown in Figure 3, the analysis revealed AUC scores above 0.9 for all malware classes, indicating good ranking ability of the classifiers.

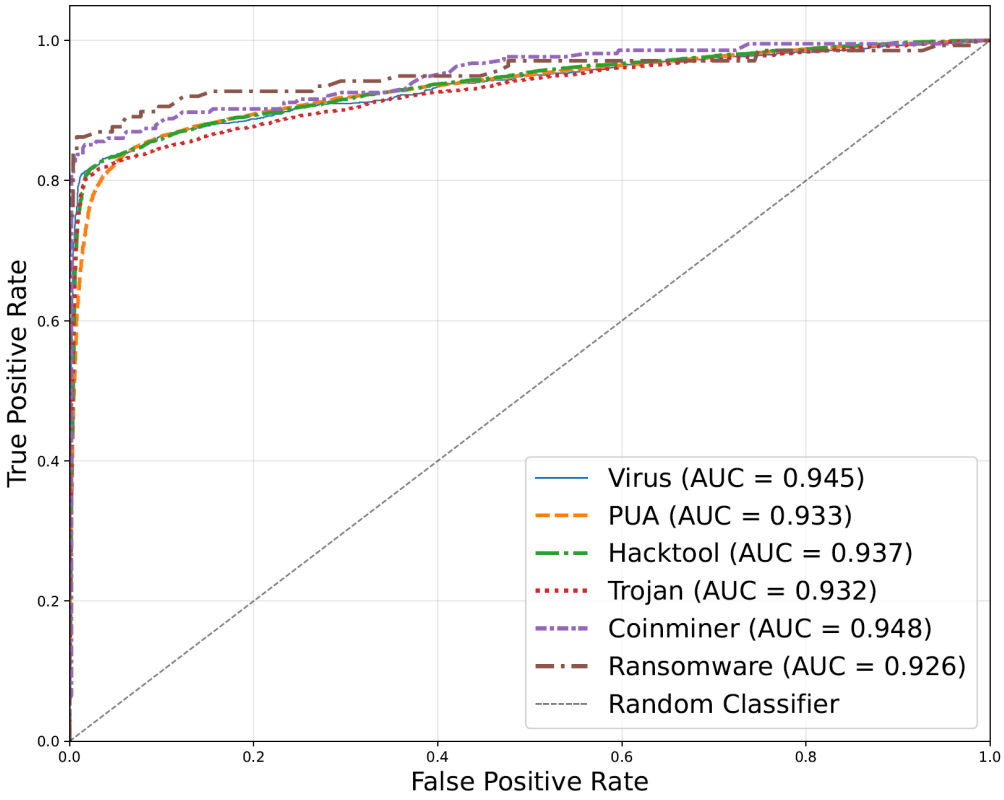


Figure 3. ROC curves for the MLC

As further optimization, dynamic thresholds were adopted. In fact, initial evaluation using the classifiers' default threshold (0.5) yielded suboptimal accuracy, despite the high AUC values. This discrepancy highlighted the need for dynamic threshold optimization, as the default threshold might not be optimal for imbalanced datasets. To address this discrepancy, Youden's J statistic method[14] was used to determine optimal classification thresholds for each binary classifier. The threshold θ was thus computed for each classifier, where TPR(θ) and FPR(θ) are the true positive and false positive rates at threshold θ, respectively:

$$\theta_{\text{optimal}} = \underset{\theta}{\text{argmax}}[\text{TPR}(\theta) - \text{FPR}(\theta)]$$

Table 6 shows the adoption of dynamic thresholds, which resulted in significant accuracy improvements across all malware classes, with F1, precision, and recall above 0.9. This indicates that greater accuracy is achieved by requiring higher confidence in classifying an endpoint as at risk. That is, a certain degree of conservatism in the judgment of the classifiers in the decision point lends to a higher overall accuracy.

	F1 score	Precision	Recall	AUC
Coinminer	0.938	0.941	0.940	0.948
Hacktool	0.924	0.927	0.925	0.937
PUA	0.904	0.904	0.904	0.933
Ransomware	0.923	0.926	0.924	0.926
Trojan	0.926	0.928	0.927	0.932
Virus	0.936	0.938	0.937	0.945

Table 6. Accuracy for the MLC: the F1 score, precision, recall, and AUC for the individual classifiers

After optimizing the thresholds, the feature importance for each malware class was computed. This analysis helped the researchers understand which user behaviors contributed most significantly to risk prediction. The feature importance for each feature was calculated using Gini impurity:[15]

$$\text{Gini}(D) = 1 - \sum_{i=1}^{c} p_i^2$$

where D represents the training set, p_i denotes the proportion of samples belonging to the malware class i, c is the total number of classes, and the sum is calculated over all classes.

To compare the feature importance across all malware classes, the importance scores were calculated by treating all different malware classes as "exposed" and classifying them collectively against the normal class. These measures reveal the relative contribution of each feature to the resulting risk score, providing interpretable insights into the behavioral patterns that help anticipate future malware outbreaks. The feature importance for the generic class "encountered malware" is depicted in Figure 4.

The results demonstrate a convergence between the MLC and ODG methodologies, despite their differing approaches. Both methods identify system- and network-reputation activities as the primary discriminative features. Notably, the feature *product version percentile rank* stands out as a significant predictor in both models. This is identified as a contributing feature in five out of six malware classes for the ODG (as shown in feature #19 in Table 5), and it is ranked among the top 10 features in the MLC, with an importance score of 0.064, as illustrated in Figure 4.

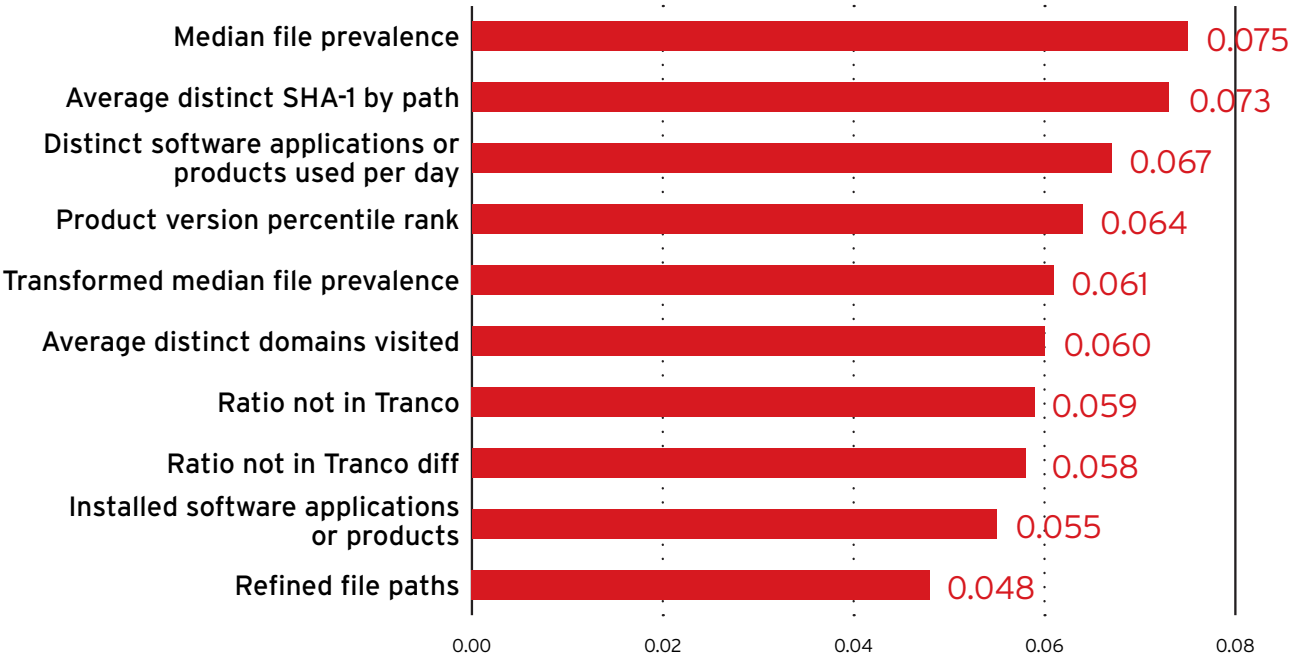


Figure 4. Feature importance for the generic class "encountered malware" (MLC)

The prominence of network behavior indicators, such as the average number of domains visited per day by a user (*average distinct domains visited*) or their lack of popularity in whitelists (*ratio not in Tranco*), underscores the importance of incorporating users' web behavior into malware risk quantification. With these features, the behavior of malware operators who tend to leverage unconventional internet facilities for their campaigns is captured.

Forecasting Future Outbreaks at Scale

This study now proceeds to estimate the risk of future malware outbreaks for new endpoints. In this phase, the approach introduced in this study was deployed in a real-world setting and utilized it to assess the risk of all endpoints collected through TrendAI™ telemetry over a one-month period (January 2025).

In total, 10,680,183 endpoints across 217 countries and 822 organizations were analyzed. The majority of these machines (73.48%) were situated in very large enterprises (VLEs, organizations with an average of over 10,000 deployed endpoints) consisting of networks used for business purposes, including regular desktop computers for day-to-day office tasks, high-end machines for specialized operations, and servers.

As presented in the overview, the estimation of the risk (on a scale between 0 and 100) consisted of a combination of a statistical estimator (odds ratios representing the association between an exposure and the outcome) and a multi-label classifier. These two approaches were combined to generate the final risk score $R = \{R_1, R_2, \ldots R_i\}$ for an endpoint e as an average of the single scores $R1$ and $R2$ (see Figure 1).

The system processing endpoint data collected in real time was left for one month and then drew cumulative results. The risks associated with various classes of malware are detailed in Table 7 and visually represented through a heatmap in Figure 5.

	Estimated risk R_i				
Malware class	**0-20**	**21-40**	**41-60**	**61-80**	**81-100**
Coinminer	45.91	44.75	8.42	0.88	0.04
Hacktool	23.37	40.55	22.21	13.08	0.79
PUA	27.82	30.85	18.82	21.60	0.91
Ransomware	32.53	39.08	20.26	7.79	0.34
Trojan	21.82	40.84	21.83	14.85	0.66
Virus	38.23	43.85	13.80	3.99	0.13
Overall risk R	31.61	39.99	17.56	10.36	0.48

Table 7. Overall risk estimation for the 10.7 million endpoints included in this study

	0-20	21-40	41-60	61-80	81-100
Coinminer	4,903,634	4,779,539	899,597	94,473	2,940
Hacktool	2,496,191	4,330,723	2,372,166	1,396,820	84,283
PUA	2,970,720	3,294,737	2,009,752	2,306,950	98,024
Ransomware	3,474,606	4,173,757	2,163,767	831,836	36,217
Trojan	2,330,561	4,362,174	2,331,336	1,586,528	69,584
Virus	4,082,714	4,683,472	1,473,422	426,139	14,436

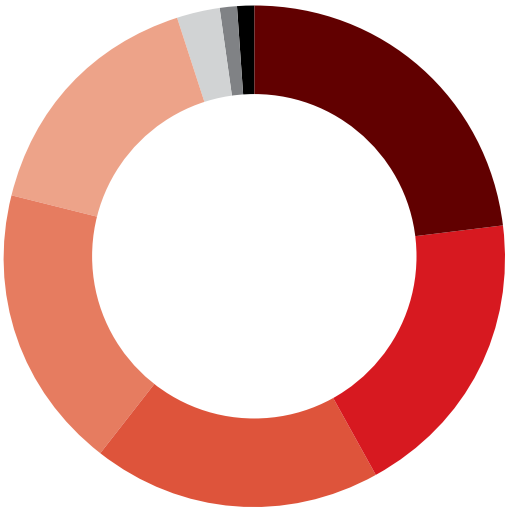
Figure 5. Risk estimation per malware class, along with the associated number of endpoints for each risk range and class

One of the first general observations after the cumulative results were collected was that 31.61% of the analyzed population falls within the 0-20 risk range. This result supports this study's initial assumption that most of the machines in the study are part of the networks of enterprises that utilize security solutions, primarily the anti-malware platform from where TrendAI™ collects its telemetry data. Furthermore, it is reasonable to assume that these organizations are mandated by regulatory frameworks such as NIS2[16] or NIST[17] to enforce corporate security policies, including web filtering and application whitelisting, which inherently reduce the overall risk a priori.

Over half (57.55%) of the machines are at medium risk (21-40 and 41-60 ranges), and 10.36% and 0.48% fall within the high 61-80 and 81-100 risk ranges, respectively. This highlights that while extreme risk is rare, the majority of users operate in a vulnerable middle zone that should still work on improving their security posture by maximizing proactive security platforms and investing in user behavior training.

When examining the different malware classes predicted by the new approach that this study tested and proposes, PUA emerges as the riskiest class in the ecosystem. This is attributable to the widespread presence of unwanted programs in all their forms, particularly on dubious websites that offer malicious applications disguised as legitimate programs, such as streaming websites providing applications for watching copyrighted content for free. This observation is further confirmed by Table 5, which shows PUA as the leading malware class when it comes to the highest prevalence of significant odds ratios (marked with a single or double *).

Conversely, coinminers are a very "specialized" form of malware that leverages the computational resources of unaware users for mining cryptocurrencies. This study reveals that the risk of encountering one of these is the lowest when compared with other classes. However, when examining the features contributing to the risk of future coinminer incidents as reported in Figure 6, visiting peer-to-peer websites (for example, those hijacked with crypto scams) or business sites (such as those offering crypto-like applications) significantly put a user at risk. This is also confirmed by all other features that capture the high number of software applications downloaded, installed, and executed by the user on the machine.



Refined file paths	23.22%
Business sites visited per day	18.99%
Product version percentile rank	18.51%
Distinct software applications or products used per day	18.32%
Peer-to-peer sites visited per day	16.20%
Malicious sites visited per day	2.62%
Suspicious and dangerous sites visited per day	1.07%
Presence of shady traffic	1.07%

Figure 6. Contributing features for the "coinminer" malware class

Next, the endpoints identified as highly exposed to malware outbreaks (with scores of 81–100 in Figure 5) were analyzed to understand their usage within corporate networks. The goal was to validate the researchers' initial assumptions that the way in which a machine is utilized is a main factor for falling victim to a specific class of malware.

For each malware class, 500 high-risk endpoints were randomly sampled, from whose system activity logs a list of installed applications, such as engineering software, enterprise resource planning (ERP) platforms, and games, were extracted. This list was further refined by excluding DLL files, drivers, and known OS applications, and removing redundant product names to improve computational performance.

This curated dataset was fed to OpenAI GPT-4o,[18] which was used to identify the intended use of each endpoint (e.g., computer graphics workstation, database server, or regular office endpoint). For this to work, a system prompt aimed at enabling the large language model (LLM) to categorize the endpoint was designed.

The results of the analysis are reported in Figure 7. Endpoints holding valuable business information critical to the operation of an organization (such as ERP or financial systems) are more susceptible to ransomware attacks, because they represent attractive targets for ransom demands. Meanwhile, coinminers tend to spread indiscriminately across the largest number of enterprise machines, which normally consist of regular desktop endpoints used for office and generic tasks. In this case, the malware authors are likely primarily interested in leveraging the largest number of machines for distributed computation of cryptocurrencies. PUAs are found targeting machines that run gaming

software, most likely because their users (gamers) are used to downloading a large variety of games, including untrusted software. This effectively shows that the way in which a user utilizes the endpoint is a main factor for categorizing the risk of future malware outbreaks.

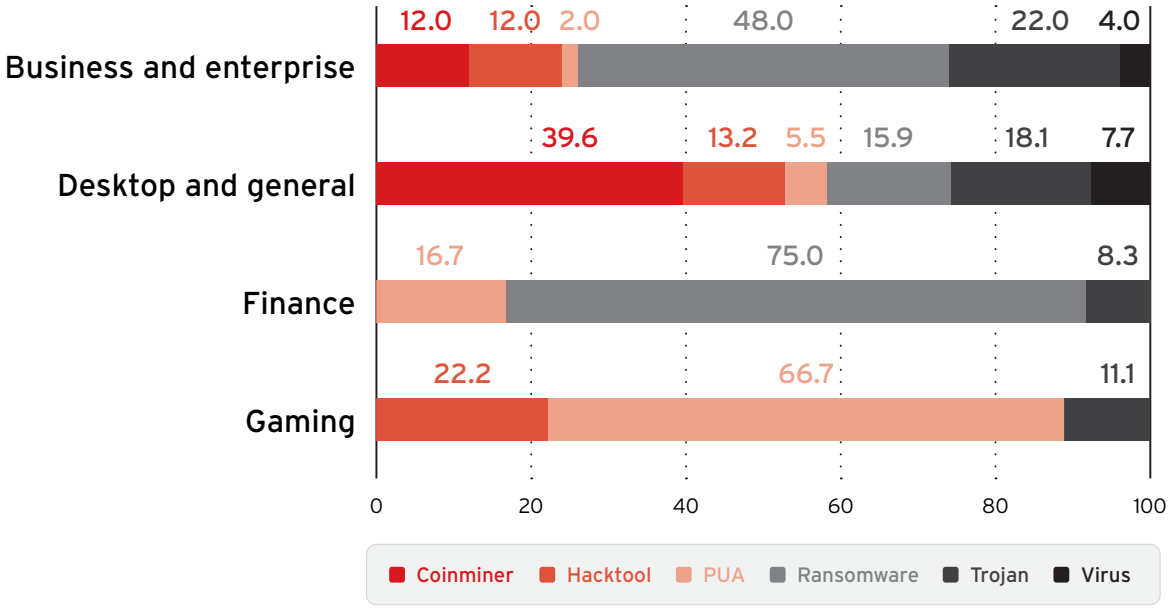


Figure 7. Number of endpoints at very high risk, categorized by endpoint usage (values expressed in percentages of the total 10.7 million endpoints analyzed)

This new approach is presently rendered in practice: A dashboard operationalizes the approach proposed in this study by visualizing the aggregated risk scores per country and class of malware using a world map, as shown in Figure 8. An analyst can interact with the map to expand on different regions or use the search functionality to look up specific endpoints, networks, or organizations. Other general information, such as the number of endpoints processed by the system, the associated organizations, and the endpoints with the highest risks, is also presented.

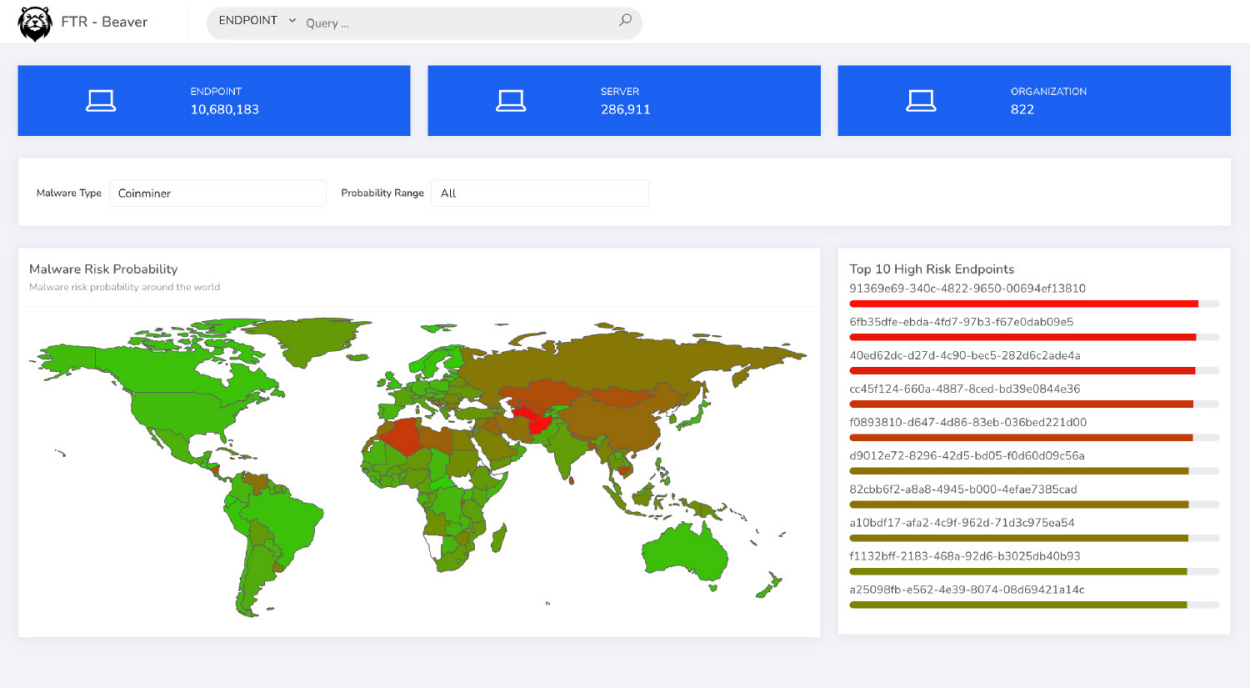


Figure 8. Screenshot of the dashboard operationalizing the new approach introduced in this study. This specific screenshot shows the risk distribution for the "coinminer" malware class.

Meanwhile, Figures 9 and 10 show screenshots from the dashboard that present a risk estimation breakdown for the specific endpoint *c46cecb2-5b8f-4635-85fd-6ea02406d0c2*. Figure 9 depicts the estimated risk for the six malware classes considered in this work and Figure 10 summarizes the behaviors contributing to the risk. These visualizations help analysts assess the factors that pose risk to the organization to prevent possible future incidents.

ENDPOINT ID : c46cecb2-5b8f-4635-85fd-6ea02406d0c2

Features Date: **2024-12-01 ~ 2024-12-31** | Operating System: **6.2** | Product: **Apex One** | Country: **United States**

Overall risk estimation: **75.24%** View risk behavior profile Recommendation

Malware Type	Probability
Ransom	96.47%
PUA	91.94%
Trojan	86.71%
Hacktool	85.56%
Virus	64.2%
Coinminer	26.57%

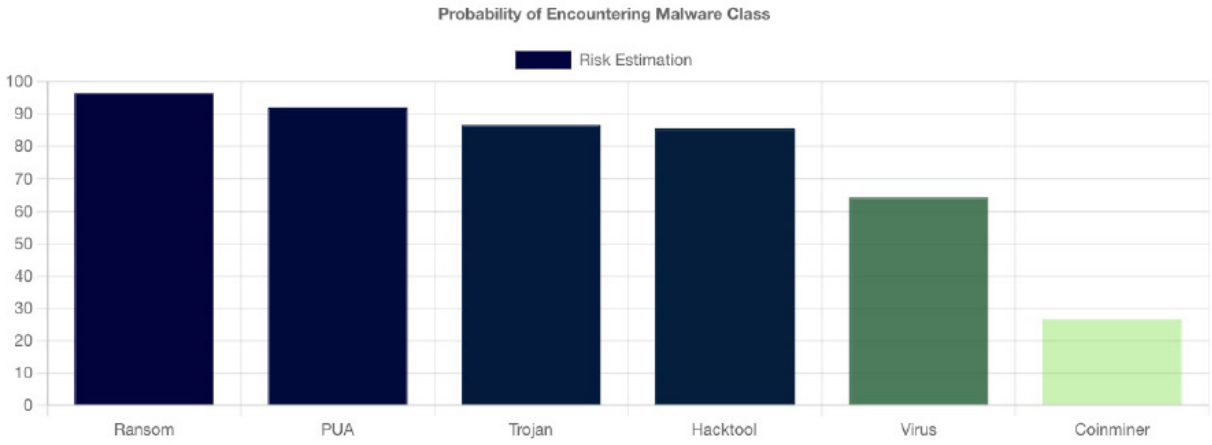


Figure 9. Estimated risk score per malware class

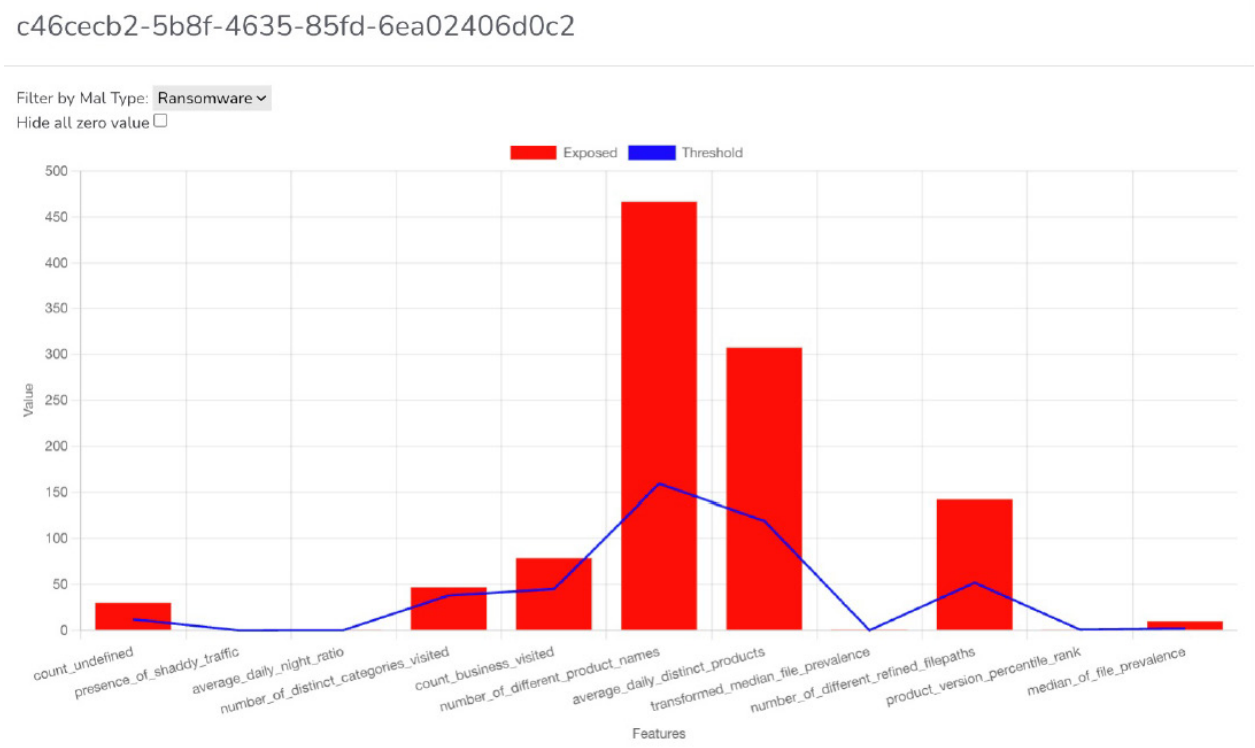


Figure 10. Contributing features for the "ransomware" malware class

Proactive Defense and Protection With TrendAI™

The limitations of reactive security approaches are well established. This study demonstrates that anticipating and mitigating risk before an incident occurs is not only possible, but also critical for effective defense.

TrendAI Vision One™[19] applies key principles to operational environments. TrendAI Vision One™ Cyber Risk Exposure Management (CREM)[20] and TrendAI Vision One™ Cyber Risk Quantification (CRQ)[21] use telemetry and analytics to assess endpoint, network, and user activity, which in turn enable the estimation of future malware outbreak probabilities and support risk-driven decision-making.

TrendAI™'s Web Reputation Service (WRS)[22] is a core component of TrendAI Vision One™ Zero Trust Secure Access (ZTSA).[23] WRS provides real-time URL filtering and threat protection as part of ZTSA's Secure Web Gateway functionality. By blocking access to dangerous and unclassified sites, WRS addresses the behavioral factors identified in this study's risk models and enforces risk-based access policies in line with zero-trust principles.

Endpoint security features, such as behavior-based ransomware detection, virtual patching, and user behavior analytics, provide additional layers of defense that adapt to changing threat conditions. The cross-layer detection and response (XDR)[24] capabilities of TrendAI Vision One™ correlate signals from endpoints, email, cloud, and network sources to identify complex attacks that might evade single-point solutions. Managed XDR services and broad ecosystem integrations support rapid onboarding, unified visibility, and accelerated response.

The research presented in this paper directly contributes to the improvement of the Cyber Risk Index (CRI)[25] within CREM, as shown in Figure 11. By employing odds-ratio analysis and multi-label classification, explainable risk scores for specific malware classes based on observed behaviors are generated. These scores are integrated into the CRI, allowing organizations to identify endpoints and users with elevated risk profiles and to prioritize controls accordingly.

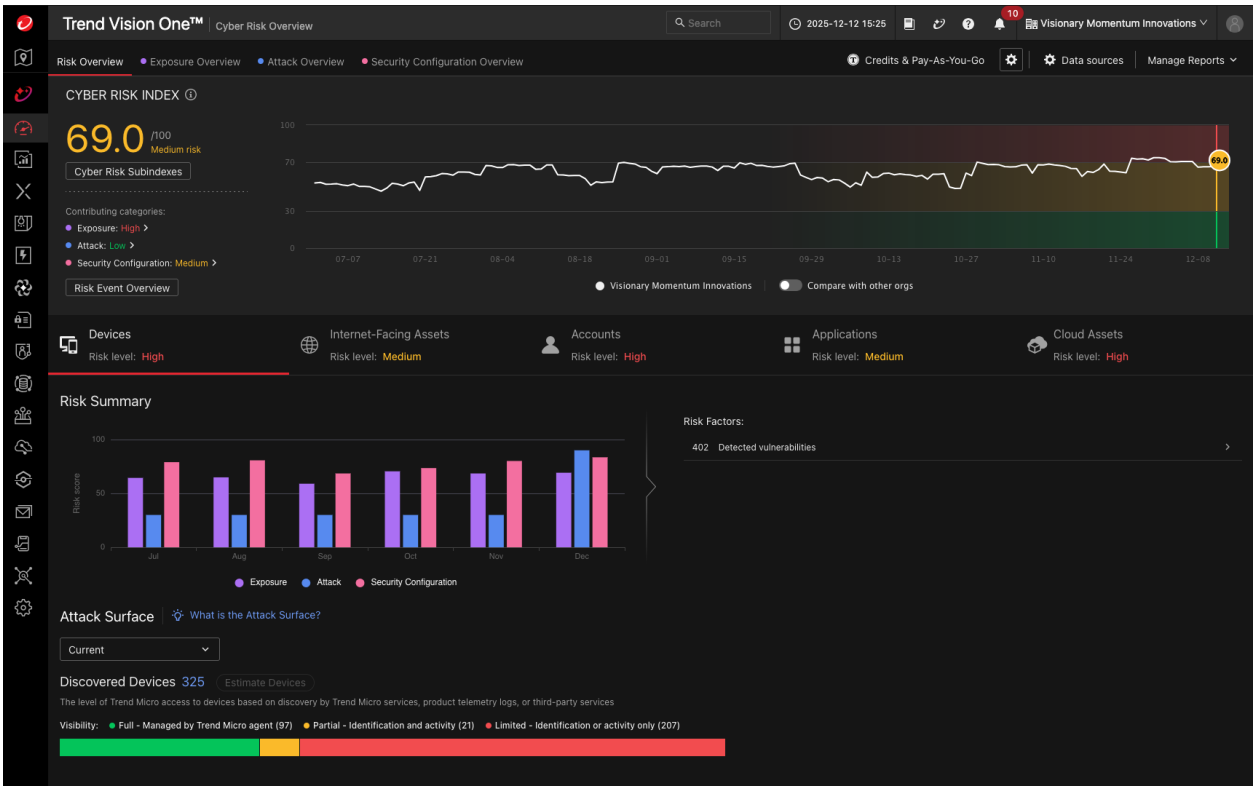


Figure 11. Visualization of the Cyber Risk Index (CRI) in TrendAI Vision One™

By combining predictive risk modeling with automated protection, TrendAI™ enables organizations to quantify, manage, and reduce cyber risk. The methodologies described in this research are actively informing the evolution of CREM and the CRI, providing a foundation for proactive security strategies that address the full spectrum of malware threats.

Conclusion

This work advances what can now be considered previous limitations of traditional reactive cybersecurity solutions by exploring the link between user behavior and potential malware outbreaks. It introduces a novel approach that provides security leaders with the ability to anticipate threats rather than react to incidents.

Findings from this study confirm that risk is highly specific as cybercriminals operate malware campaigns in different forms and with different intentions based on their need for specific types of victims and business models. A user's behavior, both on the machine and over the internet, is then a primary factor for driving infection risk estimation.

This approach has proved useful for security analysts by providing both forecasts of future compromises and insights into the behavioral drivers that lead to them. This capability represents a significant strategic advancement, allowing organizations to operationalize security resources and enforce policies preemptively.

This work drives TrendAI™ to further develop and improve solutions that enable organizations to stay ahead of cybercriminals, such as predicting the risk of future incidents across different categories of threats, which is a key component of its CREM portfolio.

Appendix

Ethical Considerations

As with prior research in this field,[26, 27, 28] the researchers behind this work believe that realistic experiments are essential for conducting reliable research, such as studies in real-world deployments. Aware of this aspect, this study prioritized user privacy and the sensitivity of the collected data. Specifically, it excluded endpoint identifiers (such as IP addresses) from its datasets and anonymized or removed any information that could reveal user identity. Users had given explicit opt-in consent for the collection of telemetry data, intended for post-sale support, threat intelligence, and additional product improvement research. There was no correlation among datasets, for example, from outside the company, which could assist in deanonymizing individuals. The datasets were accessible only within the company's secure environment to prevent unauthorized data extraction, and aggregated data was accessed by researchers through a secure repository requiring individual credentials.

Related Work

Over the past few years, several works have been conducted on identifying factors that contribute to security risks.

Woods et al.[29] presented a systematization of knowledge on quantifying cyber risk. Their work synthesized existing approaches and methodologies for measuring and modeling cybersecurity risk, aiming to provide a structured understanding of the challenges and advancements in this complex field. Their paper discussed various metrics, frameworks, and data sources used to quantify cyber risk, highlighting their strengths, limitations, and applicability in different contexts.

Lévesque et al. conducted two related clinical trials investigating human factors in home cybersecurity. Initially, in 2017,[30] they explored how user characteristics and behaviors contributed to malware risk in home networks. Building on this, their 2018 study[31] involved 50 home users, assessing antivirus performance alongside user traits like technical proficiency, browser habits, and security alert interaction to determine their influence on malware encounter risk, even with antivirus software present. Both studies highlight the critical interplay between human factors and technological defenses in shaping domestic cybersecurity.

Canali et al.[32] explored the effectiveness of predicting security risks based on users' web browsing behaviors, with data collected by an antivirus vendor. Their research aimed to determine whether it was feasible to identify users who were at a higher risk of falling victim to web-based attacks by analyzing their browsing patterns. Their study utilized a large dataset of anonymized user behavior collected by the antivirus vendor, which included details such as the types of websites visited, frequency of visits, and the user's engagement with potentially risky sites.

Yen et al.[33] investigated security risks within a large enterprise, revealing two critical insights. They found geographical variations in compromise risk, indicating location's influence. Moreover, devices faced an elevated malware risk when operating outside the enterprise's network, underscoring the protective role of organizational security perimeters.

Sanchez-Rola et al.[34] investigated the security implications associated with mouse clicks on web links. The focus of their investigation was to understand how users could be misled about the destinations of links they clicked on websites. This misleading information could significantly elevate the risks users faced, particularly in relation to security threats such as phishing attacks, where users were directed to malicious sites under false pretenses.

Thonnard et al.[35] demonstrated that the risk of encountering targeted attacks was not uniform across all industries. Their findings indicated that certain sectors were inherently more vulnerable to these sophisticated and deliberate breaches due to factors such as the value of their data, their role in critical infrastructure, or their involvement in sensitive technologies. This heightened risk necessitates a tailored approach to cybersecurity, recognizing that a one-size-fits-all defense strategy is insufficient for protecting industries facing specialized threats.

Allodi et al.[36] explored the various factors leading to vulnerability exploitation. Utilizing a case-control methodology, they found that simply relying on a high Common Vulnerability Scoring System (CVSS) score for patching offered minimal risk reduction. In contrast, the existence of a public exploit was a significantly better indicator of potential exploitation, while the presence of an exploit on cybercrime black markets proved to be the most reliable factor for predicting actual exploitation in the wild. This suggested that monitoring black markets for exploit availability could lead to more effective and economically sensible patching strategies than prioritizing solely based on CVSS scores.

Vasek et al.[37] identified specific web server characteristics that increased their likelihood of being compromised for phishing campaigns. Their research pointed to factors like the content management system (CMS) used, the web server software, and whether that software was outdated. For example, servers running popular CMS platforms like WordPress and Joomla, along with Apache and Nginx, showed higher susceptibility.

In a similar work, Di Tizio et al.[38] conducted a case-control study to identify characteristics of websites that increased their likelihood of being compromised for cryptojacking campaigns. Their preliminary analysis indicated associations between certain website features and cryptojacking.

Tajalizadehkhoob et al.[39] investigated how web security features and patching practices influenced compromise rates in shared hosting environments. Their analysis identified that webmaster and web application security efforts, including patching, significantly reduced phishing and malware incidents.

Fang et al.[40] developed a comprehensive model to assess and quantify the risk of enterprise data breaches. Their work focused on understanding the various factors that contributed to breach likelihood and impact, moving beyond qualitative assessments to provide a more data-driven approach. The model considered aspects such as an enterprise's security posture, the value of its assets, the nature of potential threats, and the effectiveness of implemented security controls.

Bilge et al.[41] developed a predictive model to identify enterprise machines at high risk of malware infection. Their approach leveraged a specific set of binary-related features, such as characteristics derived from executable files, to

determine the likelihood of a machine being compromised. By analyzing these low-level attributes, their model aimed to provide early warnings and enable proactive intervention to prevent widespread infections within an enterprise network.

In a similar work, Ovelgönne et al.[42] utilized the Worldwide Intelligence Network Environment (WINE) dataset to analyze the risk of malware attacks. Their research focused on identifying correlations between malware encounter rates and both user categories (e.g., gamers, software developers) and user system behaviors (e.g., number of binaries, binary prevalence). This allowed the researchers to measure how different user profiles and their digital habits influenced their susceptibility to malware.

Research into risk prediction extends to mobile devices, with studies by Dambra et al.[43] and Sharif et al.[44] Dambra et al. focused on Android applications, analyzing the risk of malware and PUAs based on user profiles, considering factors like application volume, diversity, and geographic features. Complementing this, Sharif et al. developed a system that predicted a user's exposure to malicious content solely by analyzing their HTTP traffic, allowing for proactive identification of risky behaviors.

Dambra et al.[45] investigated the varying incidence of malware classes (e.g., PUA, adware, and ransomware) and families between consumer and enterprise environments. Their research revealed distinct patterns of malware encounters in these two user groups. They employed a generalized linear model (GLM) to quantify the risk of these encounters based on several key characteristics, including the number of active days, geographical location, the count of installed software vendors, and file-request volume. This approach provided a granular understanding of how different factors contributed to malware risk across diverse user contexts.

The research most closely related to this work was conducted by Meschini et al.,[46] who performed a case-control study to understand how user behaviors within organizations influenced the risk of malware encounters, identifying specific network and system behaviors that increased the odds of encountering various malware. This work builds on this previous study, significantly extending to different directions, including a system that automates the estimation of future risks and malware outbreaks at scale. It leverages a wider set of features and refines the estimation by coupling the component responsible for generating the odds ratios with a multi-label classifier.

Overall, while existing literature has explored the influence of user behaviors on malware risk within organizational contexts, this research advances the field in several ways. This study provides a practical, large-scale assessment of the security risks across over 10 million endpoints used by regular users globally, far exceeding the scope of previous studies.

This work introduces a novel approach that combines statistical inference and machine learning to estimate the risk of future malware outbreaks. The researchers believe that this study is the first to conduct such an extensive study that examines the combined effect of both web behaviors and system behaviors on the risks of a wide range of malware classes, including coinminer, hacktool, PUA, ransomware, trojan, and virus. This comprehensive behavioral analysis, coupled with the new approach's implementation for practical assessment, sets this work apart from prior research that focused on specific behavioral aspects or smaller datasets, or treated the different malware classes as a whole.

References

1 Trend Micro. (n.d.). *Trend Micro*. "Cyber Risk Exposure Management." Accessed Jan. 30, 2026, at: Link.

2 A. Acar et al. (2019). "An analysis of malware trends in enterprise networks," in *Proc. Of ISC-19*.

3 P. Kotzias et al. (2019). "Mind your own business: A longitudinal study of threats and vulnerabilities in enterprises," in *Proc. Of NDSS-19*.

4 Jason Firch. (May 24, 2025). *Purplesec*. "The Average Cost of Ransomware Attacks." Accessed Jan. 30, 2026, at: Link.

5 Inventive MDR Security. (November 21, 2025). *Inventive*. "EDR vs Antivirus: Why Traditional Security Isn't Enough." Accessed Jan. 30, 2026, at: Link.

6 M. Meschini et al. (2024). "A case-control study to measure behavioral risks of malware encounters in organizations," *IEEE Transactions On Information Forensics And Security*, vol. 19, pp. 9419–9432.

7 M. Meschini et al. (2024). "A case-control study to measure behavioral risks of malware encounters in organizations," *IEEE Transactions On Information Forensics And Security*, vol. 19, pp. 9419–9432

8 M. Meschini et al. (2024). "A case-control study to measure behavioral risks of malware encounters in organizations," *IEEE Transactions On Information Forensics And Security*, vol. 19, pp. 9419–9432.

9 L. Breiman. (2001). "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32

10 V. L. Pochat et al. (2019). "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. Of NDSS-19*

11 L. Bilge, Y. Han, and M. Dell'Amico (2017). "Riskteller: Predicting the risk of cyber incidents," in *Proc. Of CCS-17*.

12 Vesselin Bontchev. (n.d.). *National Laboratory of Computer Virology*. "The Naming of Computer Viruses." Accessed on January 2026, at: Link.

13 J.-B. D. Prel et al. (2009). "Confidence interval or p-value?: Part 4 of a series on evaluation of scientific publications," *Deutsches Ärzteblatt International*, vol. 106, no. 19.

14 W. J. Youden. (1950). "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35.

15 L. Breiman. (2001). "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32.

16 European Commission. (n.d.). *European Commission*. "NIS2 Directive: securing network and information systems." Accessed Jan. 30, 2026, at: Link.

17 NIST. (n.d.). *NIST*. "Cybersecurity Framework." Accessed Jan. 30, 2026, at: Link.

18 Eric Boyd. (May 13, 2024). *Microsoft*. "Introducing GPT-4o: OpenAI's new flagship multimodal model now in preview on Azure." Accessed Jan. 30, 2026, at: Link.

19 Trend Micro. (n.d.). *Trend Micro*. "Trend Vision One™." Accessed Jan. 30, 2026, at: Link.

20 Trend Micro. (n.d.). *Trend Micro*. "Cyber Risk Exposure Management." Accessed Jan. 30, 2026, at: Link.

21 Trend Micro. (n.d.). *Trend Micro*. "Cyber Risk Quantification." Accessed Jan. 30, 2026, at: Link.

22 Trend Micro. (n.d.). *Trend Micro*. "Site Safety Center." Accessed Jan. 30, 2026, at: Link.

23 Trend Micro. (n.d.). *Trend Micro*. "Zero Trust Secure Access." Accessed Jan. 30, 2026, at: Link.

24 Trend Micro. (n.d.). *Trend Micro*. "Security Operations." Accessed Jan. 30, 2026, at: Link.

25 Trend Micro. (n.d.). *Trend Micro*. "Cyber Risk Scoring." Accessed Jan. 30, 2026, at: Link.

26 M. Jakobsson and J. Ratkiewicz. (2006). "Designing ethical phishing experiments: A study of (ROT13) ronl query features," in *Proc. Of WWW-06*.

27 M. Jakobsson, P. Finn, and N. A. Johnson. (2008). "Why and how to perform fraud experiments," *IEEE Secur. Priv.*, vol. 6, no. 2.

28 T. Yen and M. K. Reiter. (2008). "Traffic aggregation for malware detection," in *Proc. Of DIMVA-08*.

29 D. W. Woods and R. Böhme. (2021). "Sok: Quantifying cyber risk," in *Proc. Of SS&p-21*.

30 F. L. Lévesque, J. M. Fernandez, and D. Batchelder. (2017). "Age and gender as independent risk factors for malware victimisation," in *Proc. Of BCS HCI-17*.

31 F. L. Lévesque et al. (2018). "Technological and human factors of malware attacks: A computer security clinical trial approach," ACM Trans. *Priv. Secur.*, vol. 21, no. 4.

32 D. Canali, L. Bilge, and D. Balzarotti. (2014). "On the effectiveness of risk prediction based on users browsing behavior," in *Proc. Of AsiaCCS-14*.

33 T. Yen et al. (2014). "An epidemiological study of malware encounters in a large enterprise," in *Proc. Of CCS-14*.

34 I. Sánchez-Rola et al. (2020). "Dirty clicks: A study of the usability and security implications of click-related behaviors on the web," in *Proc. Of WWW-20*.

35 O. Thonnard et al. (2015). "Are you at risk? Profiling organizations and individuals subject to targeted attacks," in *Proc. Of FC-15*.

36 L. Allodi and F. Massacci. (2014). "Comparing vulnerability severity and exploits using case-control studies," *ACM Trans. Inf. Syst. Secur.*, vol. 17, no. 1.

37 M. Vasek, J. Wadleigh, and T. Moore. (2016). "Hacking is not random: A case-control study of webserver-compromise risk," *IEEE Trans. Dependable Secur. Comput.*, vol. 13, no. 2.

38 G. D. Tizio and C. N. Ngo. (2020). "Are you a favorite target for cryptojacking? A case-control study on the cryptojacking ecosystem," in *Proc. Of IEEE EuroS&PW-20*.

39 S. Tajalizadehkhoob et al. (2017). "Herding vulnerable cats: A statistical approach to disentangle joint responsibility for web security in shared hosting," in *Proc. Of CCS-17*.

40 Z. Fang et al. (2021). "A framework for predicting data breach risk: Leveraging dependence to cope with sparsity," *IEEE Trans. Inf. Forensics Secur.*, vol. 16.

41 L. Bilge, Y. Han, and M. Dell'Amico (2017). "Riskteller: Predicting the risk of cyber incidents," in *Proc. Of CCS-17*.

42 M. Ovelgönne et al. (2017). "Understanding the relationship between human behavior and susceptibility to cyber attacks: A data-driven approach," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 4.

43 S. Dambra et al. (2023). "One size does not fit all: Quantifying the risk of malicious app encounters for different android user profiles," in *Proc. Of USENIX-23*.

44 M. Sharif et al. (2018). "Predicting impending exposure to malicious content from user behavior," in *Proc. Of CCS-18*.

45 S. Dambra, L. Bilge, and D. Balzarotti. (2023). "A comparison of systemic and systematic risks of malware encounters in consumer and enterprise environments," *ACM Trans. Priv. Secur.*, vol. 26, no. 2.

46 M. Meschini et al. (2024). "A case-control study to measure behavioral risks of malware encounters in organizations," *IEEE Transactions On Information Forensics And Security*, vol. 19, pp. 9419–9432.

Copyright © 2026 Trend Micro Incorporated. All rights reserved. Trend Micro, the Trend Micro logo, the t-ball logo, and TrendAI are trademarks or registered trademarks of Trend Micro Incorporated. All other company and/++or product names may be trademarks or registered trademarks of their owners. Information contained in this document is subject to change without notice. Trend Micro, the Trend Micro logo, and the t-ball logo Reg. U.S. Pat. & Tm. Off. [REPO1_Research Report_Template_A4_241223US].

TrendMicro.com