

Data-Driven Cybersecurity Risk Assessment in FinTech: A Mixed-Methods Analysis of Breach Patterns, Vulnerability Factors and Financial Impact, 2008-2025

Aahwaan Khullar

Shiv Nadar School, Faridabad, India
Email: aahwaankhullar[at]gmail.com
September 2026

Preprint note: A preprint of this paper has been submitted to SSRN (Social Science Research Network). It has not been peer-reviewed or published in any journal.

Abstract: As FinTech has grown, digital financial platforms have become major targets for cybercriminals, yet most research on FinTech cybersecurity only describes incidents without measuring patterns or risk. In this study, I built a dataset of 47 publicly reported FinTech cyber incidents from 2008 to 2025, recording the type of company, attack method, records exposed, financial loss and root causes of each, and analysed it using statistical tests and a content analysis of incident narratives. Three clear patterns emerged. First, damage is highly uneven, with just three incidents accounting for almost 70% of the 1.69 billion records exposed. Second, attacks increasingly come through outside vendors and partners. Third-party involvement rose from zero incidents before 2020 to half of all incidents in 2023–2025, and ransomware caused six of the sixteen most recent cases. Third, crypto and DeFi platforms suffered far larger losses, with a median loss of US\$290 million compared with US\$26 million for other FinTech firms. My findings suggest that FinTech cyber risk now spreads through vendors and shared systems across the ecosystem rather than through attacks on single companies. Based on this, I propose a five-stage framework that FinTech firms can use to assess and prioritise their cyber risks.

Keywords: FinTech Cybersecurity, Cyber Risk Assessment, Data Breaches, Third-Party Risk, Ransomware

1. Introduction

1) Background

Over the past two decades, the financial industry has gone through a massive digital transformation. Financial technology (FinTech) companies now provide digital wallets, payment gateways, peer-to-peer lending, neobanking, online brokerage and digital asset trading to millions of users all over the world, and they have changed both how financial services are delivered and what customers expect from them [28], [29], [34]. In India alone, platforms built on UPI, digital lending and online investing have brought a large part of the population into formal digital finance within a few years. However, the same features that make FinTech attractive, which include speed, constant connectivity, open APIs, cloud infrastructure and heavy dependence on partners, have also made the sector one of the most targeted and most exposed industries for cyberattacks.

The financial cost of this exposure is significant. According to the IBM and Ponemon Institute study, the financial sector recorded an average breach cost of US\$5.56 million in 2025, the second highest of all industries after healthcare, while the global cross-industry average was US\$4.44 million [14]. Furthermore, the Verizon Data Breach Investigations Report found that the share of breaches with third-party involvement doubled from 15% to 30% in its 2025 edition and that ransomware was present in 44% of all breaches analysed [15]. At the macro level, the International Monetary Fund has warned that cyber incidents at financial firms can cause losses large enough to threaten the stability of the financial system itself [12], [13].

2) Research Gap and Problem Definition

Earlier work in this area has largely taken the form of secondary research. Khullar [1], in *Data Breaches and Cybersecurity in Fintech – A Comprehensive Analysis*, reviewed the major categories of cyber threats to FinTech platforms, including malware, phishing, data leakages, DDoS attacks, API exploits and cloud security risks, and examined countermeasures such as end-to-end encryption, multi-factor authentication, AI-powered threat detection and blockchain technology along with their cost structures, drawing primarily on existing literature and published case evidence. That study, like most of the existing literature on FinTech cybersecurity [2], [3], [25], was qualitative in nature and relied on case studies and scholarly review. While such work is valuable for mapping the threat landscape, it cannot answer questions that require measurement, such as which attack vectors are becoming more common over time, which vulnerability factors are linked to larger breaches, and which risks a FinTech firm should prioritise first with limited resources.

On the other hand, quantitative research on data breaches, such as studies on breach size distributions [8], [9], breach costs [7], [10] and the drivers of cyber risk [11], generally uses cross-industry datasets and does not focus on the specific structure of the FinTech ecosystem, where banks, payment firms, lenders, brokers and crypto platforms are tightly linked through shared vendors and infrastructure. There is therefore a clear gap for a study that applies structured, data-driven methods specifically to FinTech incidents and turns the results into a practical risk assessment framework.

Volume 15 Issue 10, October 2026

Fully Refereed | Open Access | Double Blind Peer Reviewed Journal

www.ijsr.net

3) Objectives and Research Questions

This study aims to develop a data-driven framework for assessing cybersecurity risk in FinTech by analysing breach patterns, identifying key vulnerability factors and estimating the potential financial and operational impact of incidents. In order to achieve this aim, the study addresses four research questions:

- **RQ1:** How have the frequency and nature of FinTech cybersecurity incidents changed between 2008 and 2025?
- **RQ2:** Which attack vectors and root-cause vulnerability factors are most common, and does their distribution differ across FinTech sub-sectors?
- **RQ3:** How are the size and financial impact of FinTech breaches distributed, and which factors are associated with larger impact?
- **RQ4:** How can these findings be combined into a prioritised risk assessment that FinTech firms and regulators can use?

4) Contributions

This study makes three contributions. Firstly, it builds a structured and openly documented dataset of 47 FinTech incidents coded on the same set of variables, which can be extended by future researchers. Secondly, it combines quantitative statistical analysis with a validated text-based content analysis of incident narratives, so that numerical patterns can be explained by the underlying causes of each incident. Thirdly, it converts the results into a recency-weighted risk index and a five-stage FinTech Cyber Risk Assessment Framework, extending earlier literature-based reviews of FinTech threats [1], [2] towards evidence-based prioritisation.

5) Organisation of the Paper

The literature is reviewed in Section II. The dataset, coding scheme and analytical methods are explained in Section III. The quantitative and text analysis results are presented in Section IV, and they are discussed along with the proposed framework in Section V. Limitations and future research directions are covered in Section VI, and the paper is concluded in Section VII.

2. Literature Survey

The literature relevant to this study can be grouped into four streams: studies on the cyber threat landscape of FinTech, quantitative studies on data breach size and cost, studies that link cyber risk to financial stability and regulation, and methodological work on risk assessment and content analysis.

1) Cyber Threats in the FinTech Sector

Javaheri et al. [2], in their systematic review "Cybersecurity threats in FinTech", analysed 74 academic studies and categorised eleven distinct cyber threats to FinTech platforms, including data breaches, ransomware, phishing and insider threats, and linked them to nine defence strategies. The authors strongly argue that innovations such as blockchain, AI and digital payments increase the attack surface of FinTech firms, and that multi-layered security and user awareness are essential to prevent social engineering attacks. Olaiya et al. [3] similarly emphasised the massive need for stronger cybersecurity measures in FinTech, examining malware, DDoS, phishing and data breaches along

with the regulatory pressure created by frameworks such as GDPR and PCI DSS. Gosha [25] studied the simultaneous growth of FinTech and its cybersecurity challenges, identifying data breaches, ransomware, phishing and cloud misconfigurations as key threats and highlighting India's move towards self-regulatory organisations. Kshetri [4] analysed the causes and consequences of cybercrime in India, linking rapid digitalisation and weak cybersecurity infrastructure to rising attacks on financial institutions and their clients.

Khullar [1] reviewed these threat categories in detail through a synthesis of secondary literature, discussing real incidents such as the Bangladesh Bank SWIFT heist, the Capital One cloud breach and the Traveler ransomware attack, along with the costs of implementing key countermeasures. A common feature of this stream of research is that it is largely qualitative: threats are described and illustrated with cases, but their relative frequency, their trend over time and their link to impact are not measured.

2) Quantitative Studies on Data Breach Size and Cost

A second stream of research treats breaches as measurable events. Edwards, Hofmeyr and Forrest [8] analysed a decade of reported US breaches and found that breach sizes follow a heavy-tailed, log-normal distribution and that, contrary to popular belief, neither the size nor the frequency of breaches had increased significantly over the period studied. Wheatley, Maillart and Sornette [9] studied the largest personal data breaches and showed that their size follows an extreme, power-law-like tail, which means that a few very large events dominate total losses. Romanosky [7] examined over 12,000 cyber incidents and found that the typical cost of an incident was far lower than commonly assumed, while a small number of events were extremely costly. Eling and Wirfs [10] used operational loss data to show that cyber risk events have a heavy tail and that extreme events should be modelled separately from everyday losses. Aldasoro et al. [11] studied the drivers of cyber risk and found that the financial sector faces a higher frequency of cyber incidents but, on average, lower costs than some other sectors, which they linked to higher investment in security.

Industry reports provide the most up-to-date aggregate figures. The IBM and Ponemon Institute study reported a 2025 average breach cost of US\$5.56 million for the financial sector and identified phishing (16%) and supply-chain compromise (15%) as the two most common initial attack vectors [14]. The Verizon DBIR 2025, based on 12,195 confirmed breaches, found that credential abuse (22%) and exploitation of vulnerabilities (20%) were the leading initial access vectors, and that the human element was involved in around 60% of breaches [15]. Although these reports are extremely useful, they are cross-industry, rely on proprietary or survey data and do not publish incident-level data that can be re-analysed for the FinTech sector specifically.

3) Cyber Risk, Financial Stability and Regulation

Kopp, Kaffenberger and Wilson [12] argued that cyber risk in finance involves market failures, because firms do not bear the full cost of the risk they create for connected institutions, and that concentration in shared service providers can turn a single incident into a system-wide event. The IMF's Global

Financial Stability Report [13] reinforced this view, finding that the financial sector is highly exposed to cyber incidents and that extreme losses are becoming more severe. In Europe, ENISA's sector-specific threat landscape [26] similarly identifies ransomware, DDoS and third-party compromise as leading threats to financial institutions. Najaf, Mostafiz and Najaf [5] linked cybersecurity risk directly to the sustainability of FinTech firms and banks. In response, regulators have strengthened their frameworks: the EU's Digital Operational Resilience Act (DORA) [19] explicitly covers ICT third-party risk, the Reserve Bank of India's Master Direction on IT governance [17] sets expectations for Indian regulated entities, and India's Digital Personal Data Protection Act, 2023 [18] introduces obligations for protecting personal data.

4) Risk Assessment and Content Analysis Methods

Formal cyber risk assessment usually combines the likelihood of a threat event with its impact, as described in NIST SP 800-

30 [21], and the NIST Cybersecurity Framework 2.0 [20] adds a "Govern" function that includes supply-chain risk management. Gordon and Loeb [6] showed economically that security investment should be matched to the vulnerability and potential loss of the information being protected, which supports prioritising risks rather than protecting everything equally. For analysing textual material, qualitative content analysis provides systematic ways to code text into categories [22], [23], and inter-coder agreement measured by Cohen's kappa [30] is commonly used to check the reliability of such coding, with values above 0.80 usually interpreted as almost perfect agreement [24].

Table 1 synthesises the key studies reviewed above, along with their methods and their limitations in relation to the objectives of this study.

Table 1: Synthesis of key prior studies

Study	Focus	Method	Key finding	Limitation addressed by this study
Javaheri et al. (2024) [2]	Cyber threats in FinTech	Systematic review of 74 studies	Eleven threat types mapped to nine defence strategies	No incident-level measurement of frequency or impact
Olaiya et al. (2024) [3]	FinTech security strategies	Narrative review	Layered security, AI/ML and regulatory compliance are essential	Descriptive; no empirical testing
Khullar (2025) [1]	FinTech breaches and countermeasures	Secondary research and case studies	Malware, phishing, data leakage and DDoS as major threats; cost structures of countermeasures	No statistical analysis of incident patterns
Kshetri (2016) [4]	Cybercrime in India	Qualitative institutional analysis	Rapid digitalisation and weak infrastructure drive cybercrime	Not FinTech-specific; no incident dataset
Edwards et al. (2016) [8]	US data breaches	Statistical modelling of breach chronology	Breach sizes log-normal; no significant growth trend	Cross-industry; no sub-sector analysis
Wheatley et al. (2016) [9]	Largest personal data breaches	Extreme value statistics	Power-law-like tail dominates total losses	Cross-industry; size only
Romanosky (2016) [7]	Costs and causes of cyber incidents	Econometric analysis of 12,000+ incidents	Typical costs low, a few events extremely costly	Cross-industry; limited FinTech coverage
Eling and Wirfs (2019) [10]	Cost of cyber risk events	Operational loss data, extreme value theory	Heavy-tailed losses require separate tail modelling	Operational risk data; no attack-vector analysis
Aldasoro et al. (2022) [11]	Drivers of cyber risk	Econometric analysis	Finance faces more frequent but relatively lower-cost incidents	Sector-level aggregates; no root-cause coding
Kopp et al. (2017) [12]	Cyber risk and financial stability	Conceptual and policy analysis	Concentration in shared providers creates systemic risk	No incident-level evidence

5) Identified Gap

Based on this review, three gaps can be identified. Firstly, FinTech-specific cybersecurity research is mostly descriptive and does not quantify patterns across incidents [1], [2], [3]. Secondly, quantitative breach studies [7], [8], [9], [10], [11] do not focus on the FinTech ecosystem and its sub-sectors, such as payments, lending and crypto platforms. Thirdly, few studies combine incident-level statistics with text analysis of incident causes and then convert the results into a practical risk prioritisation tool. This study attempts to address all three gaps.

6) Hypothesis Development

Based on the literature reviewed above, six hypotheses were developed in relation to the research questions.

The Verizon DBIR [15] and the IBM study [14] report a sharp rise in supply-chain and third-party compromise, while Kopp et al. [12] argue that dependence on shared service providers concentrates cyber risk. FinTech firms, which rely on partner banks, cloud providers and outsourced services by design, are expected to reflect this trend strongly.

- **H1:** The share of FinTech incidents involving third parties has increased over time.

Popular accounts often suggest that breaches are becoming larger, whereas Edwards et al. [8] found no significant growth in breach size in cross-industry data.

- **H2:** The size of FinTech breaches, measured by records exposed, has increased over time.
- Prior quantitative studies consistently report heavy-tailed breach sizes and losses [7], [8], [9], [10].
- **H3:** The impact of FinTech breaches follows a heavy-tailed, approximately log-normal distribution.
- Since FinTech sub-sectors differ in their technology stacks and in what attackers can steal, from customer records to directly transferable digital assets [2], [16], both the routes of attack and the resulting losses are expected to differ.
- **H4:** The distribution of primary attack vectors differs across FinTech sub-sectors.
- **H5:** Crypto and DeFi incidents result in larger financial losses than incidents in other FinTech sub-sectors.

- Finally, the threat literature [1], [2], [3] implies that certain weaknesses, particularly technical misconfigurations and weak authentication, lead to more severe breaches.
- H6:** Root-cause vulnerability factors are significantly associated with the number of records exposed.
- H1 and H2 address RQ1, H4 addresses RQ2, and H3, H5 and H6 address RQ3. RQ4 is addressed through the risk index and the framework derived from the tested results.

3. Methodology

1) Research Design

This study follows a mixed-methods, explanatory design. The quantitative strand measures how FinTech incidents are

distributed across time, sub-sectors, attack vectors and impact, while the text-based strand codes the written narrative of each incident into root-cause vulnerability factors. The two strands are then integrated into a likelihood $\times$ impact risk index, following the logic of NIST SP 800-30 [21]. All analysis was carried out in Python using pandas, NumPy, SciPy and scikit-learn.

The analytical workflow is organised into six sequential stages, as shown in Fig. 1. Each stage produces a defined output that becomes the input for the next stage, which ensures that every reported figure can be traced back to the coded dataset.

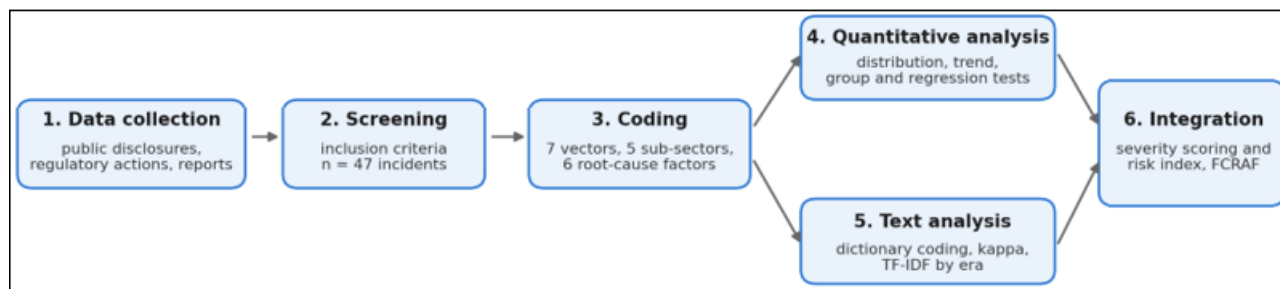


Figure 1: Research design and data analysis pipeline

In Stage 1, candidate incidents were identified from public sources. In Stage 2, the inclusion criteria in Section III-B were applied. In Stage 3, each incident was coded on the variables in Table 2 and given a structured narrative. Stages 4 and 5 were run in parallel on the same coded dataset, and Stage 6 combined their outputs into the severity score, the risk index and the proposed framework.

2) Data Collection and Inclusion Criteria

Since no open incident-level database exists for the FinTech sector specifically, a structured dataset was compiled by the author from publicly available sources, which include company breach notifications and annual reports, regulatory enforcement actions and settlements, official statements from affected firms, and established cybersecurity and financial news reporting. This approach follows earlier breach studies that relied on public breach chronologies [8], [27]. An incident was included if it met all of the following criteria:

- The affected organisation provides financial services through digital channels, or is a technology provider whose compromise directly exposed financial data or

funds (banks and credit unions, payment firms and neobanks, lenders, mortgage servicers and credit bureaus, brokers and insurers, and crypto exchanges and DeFi protocols).

- The incident was a confirmed or credibly reported security breach involving unauthorised access to data, systems or funds, and not only a service outage.
- At least one impact measure (records exposed or financial loss) or a clearly described attack method was publicly reported.
- The incident occurred between 2008 and 2025.

The final dataset contains 47 incidents across 14 countries and five regions. Three incidents (MobiKwik 2021, TransUnion South Africa 2022 and Santander 2024) have record counts that were claimed by attackers or disputed by the company; these were flagged and a sensitivity analysis was run without them. The full dataset is provided in Appendix A.

3) Variables and Coding

Each incident was coded on the variables listed in Table 2.

Table 2: Variables and coding scheme

Variable	Type	Description
Year	Numeric	Year the incident occurred or was disclosed
Sub-sector	Categorical (5)	Banks & credit unions; Payments & neobanks; Lending, mortgage & credit; Brokerage & insurance; Crypto & DeFi
Region	Categorical (5)	North America; Asia-Pacific; Europe; Africa & Middle East; Global/Offshore
Primary attack vector	Categorical (7)	Ransomware; Intrusion & malware; Credential & access compromise; Misconfiguration & software flaw; Human manipulation & insider; Third-party & supply chain; Crypto key & wallet compromise
Records exposed	Numeric (millions)	Number of customer records or accounts affected, where reported (n = 31)
Financial loss	Numeric (US\$ millions)	Funds stolen, ransom paid, settlement or regulatory penalty as publicly reported (n = 27)
Root-cause factors	Binary (6)	Human factor; Third-party involvement; Weak authentication or credential control; Misconfiguration or software flaw; Cryptographic key compromise; Encryption failure
Narrative	Text	A structured 1–2 sentence description of how the incident happened

The primary attack vector records the main route of compromise, while the root-cause factors are not mutually exclusive, since one incident can involve several weaknesses at once (for example, a phishing attack on a vendor employee that leads to key compromise). For analysis, records and losses were transformed using log10 because both variables are highly skewed.

4) Text Analysis: Content Analysis of Incident Narratives

The text strand used directed (deductive) content analysis [22], [23]. A codebook of six root-cause factors was derived from the threat categories in the literature [1], [2], [15]. Each narrative was first coded manually by the author, following a corpus-based coding approach comparable to that applied in earlier computational analysis of stakeholder framing in the Swachh Bharat Mission [35]. In order to make the coding reproducible and check its consistency, a dictionary-based classifier was then built in which each factor is represented by a set of keyword patterns (for example, "phish", "social engineering", "bribed" and "former employee" for the human factor, and "vendor", "service provider" and "outsourced" for third-party involvement; the full lexicon is given in Appendix B). Agreement between manual and automated coding was measured using Cohen's kappa [30] for each factor and for all factors pooled. In addition to this, term frequency-inverse document frequency (TF-IDF) weighting [31] with unigrams and bigrams was applied to compare the dominant vocabulary of incident narratives before 2020 and from 2020 onwards, along with the share of narratives containing key threat terms in each era.

5) Statistical Analysis

The following methods were applied:

- **Descriptive statistics and distribution tests:** mean, median, interquartile range, skewness and the share of total impact held by the top 10% of incidents; Shapiro-Wilk tests on raw and log-transformed values to test for a log-normal distribution.
- **Trend analysis (RQ1):** Spearman rank correlation between year and log impact, and a logistic regression of third-party involvement on year, reported as an odds ratio.
- **Group comparisons (RQ2, RQ3):** the Kruskal-Wallis test [32] for differences in log records across attack vectors, the Mann-Whitney U test [33] for crypto versus non-crypto losses and for incidents with and without each root-cause factor (with rank-biserial effect sizes), and Fisher's exact test for specific 2×2 associations.
- **Sub-sector $\times$ vector association:** since every expected cell count in the 5×7 table was below five, the standard chi-square approximation is not valid, so a permutation chi-square test with 10,000 random permutations was used, with Cramér's V as the effect size.
- **Multivariable model:** an ordinary least squares regression of log10 records on year and four root-cause factors, using heteroskedasticity-robust (HC3) standard errors.

A significance level of 0.05 was used throughout. Given the modest sample size, results are interpreted with attention to effect sizes and not only p-values.

6) Risk Index Construction (RQ4)

In order to combine likelihood and impact, each incident was first given a severity score from 1 to 5 using the ordinal bands

in Table 3, and where both records and loss were available the higher band was taken:

Table 3: Severity scoring bands

Severity	Records exposed	Financial loss
1	< 0.1 million	< US\$10 million
2	0.1–1 million	US\$10–50 million
3	1–10 million	US\$50–200 million
4	10–100 million	US\$200–500 million
5	> 100 million	> US\$500 million

Likelihood was measured as the recency-weighted share of incidents in each category, so that recent patterns count more towards current risk. Incidents from 2008–2019 received a weight of 1.0, from 2020–2022 a weight of 1.5, and from 2023–2025 a weight of 2.0. The weighted shares were rescaled to a 1–5 range, and the risk score was calculated as:

$$\text{Risk}_c = L_c \times I_c, \quad L_c = 1 + 4 \frac{s_c - s_{\min}}{s_{\max} - s_{\min}}, \quad s_c = \frac{\sum_i \in c w_i}{\sum_i w_i}$$

where c is an attack vector (or sub-sector), w is the recency weight of incident i , and I is the mean severity of incidents in category c . The same index was calculated for sub-sectors.

The procedure for computing the risk index is summarised in Algorithm 1.

Algorithm 1: Recency-weighted risk index

- For each incident i , assign a records band and a loss band using Table 3, and set the severity $S(i)$ as the maximum of the available bands.
- Assign a recency weight $w(i)$ of 1.0 (2008–2019), 1.5 (2020–2022) or 2.0 (2023–2025).
- For each category c , compute the weighted share $s(c)$ as the sum of $w(i)$ for incidents in c , divided by the sum of all weights.
- Rescale the shares to a likelihood score $L(c)$ between 1 and 5 using min-max scaling.
- Compute the impact score $I(c)$ as the mean of $S(i)$ over incidents in c that have at least one impact measure.
- Compute $\text{Risk}(c) = L(c) \times I(c)$ and rank the categories in descending order.
- Repeat Steps 2–6 under alternative weighting schemes and compare the rankings (Section III-H).

7) Formal Specification of the Statistical Models

This subsection states each model formally, so that the analysis can be reproduced exactly.

Transformation. Records exposed $R(i)$ (in units) and financial loss $F(i)$ (in US\$) were log-transformed before inferential testing because both variables are strongly right-skewed:

$$r_i = \log_{10}(R_i), \quad f_i = \log_{10}(F_i)$$

Trend in breach size (H2). The monotonic association between year and log impact was measured using Spearman's rank correlation, where $d(i)$ is the difference between the ranks of the two variables for incident i :

$$\rho = 1 - \frac{\sum_{i=1}^n d_i^2}{n(n^2-1)}$$

Trend in third-party involvement (H1). A binary logistic regression was estimated by maximum likelihood, with

standard errors taken from the inverse of the observed information matrix:

$$\log\left(\frac{P(\text{TP}_i = 1)}{1 - P(\text{TP}_i = 1)}\right) = \beta_0 + \beta_1 (\text{Year}_i - 2008), \quad \text{OR} = e^{\beta_1}$$

Group differences (H5, H6). For two groups of sizes n_1 and n_2 , the Mann–Whitney U statistic was used, with the rank-biserial correlation as the effect size:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T_1, \quad r_{rb} = 1 - \frac{2U}{n_1 n_2}$$

where T_1 is the rank sum of the first group. For more than two groups, the Kruskal–Wallis statistic was used, where N is the total sample size, k the number of groups and $T(j)$ the rank sum of group j :

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} - 3(N+1)$$

The difference in median loss between crypto and non-crypto incidents was additionally estimated with a percentile bootstrap confidence interval using 10,000 resamples drawn with replacement within each group.

Association between sub-sector and vector (H4). The chi-square statistic of the 5×7 contingency table was compared against its permutation distribution, obtained by randomly shuffling vector labels $B = 10,000$ times:

$$p_{\text{perm}} = \frac{1}{B} \sum_{b=1}^B \mathbb{1}\left\{\chi^2_{\text{obs}} \geq \chi^2_{\text{obs}}\right\} \quad \text{OR} = \sqrt{\frac{\chi^2_{\text{obs}}}{N(\min(r,c) - 1)}}$$

Multivariable model (H6). Log records were regressed on year and four root-cause indicators:

$$\text{r}_i = \beta_0 + \beta_1 \text{Year}_i + \beta_2 H_i + \beta_3 \text{TP}_i + \beta_4 \text{WA}_i + \beta_5 \text{MS}_i + \varepsilon_i$$

Because the residual variance is unlikely to be constant in a heavy-tailed sample, heteroskedasticity-consistent HC3 standard errors were used, where X is the design matrix, $e(i)$ the residual and $h(ii)$ the leverage of incident i :

$$\widehat{\text{Var}}(\hat{\beta}) = (X^{\text{top}} X)^{-1} X^{\text{top}} \text{diag}\left(\frac{e_i^2}{(1 - h_{ii})^2}\right) X (X^{\text{top}} X)^{-1}$$

Coding reliability. Agreement between manual and dictionary-based coding was measured with Cohen's kappa, where $p(o)$ is the observed agreement and $p(e)$ the agreement expected by chance:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Term weighting. For the era comparison, all narratives within an era were concatenated into one document, giving $N = 2$ documents. Each term t in document d was weighted using sublinear term frequency and smoothed inverse document frequency, and each document vector was L2-normalised:

$$w_{\{t,d\}} = \left(1 + \ln \text{tf}_{\{t,d\}}\right) \times \left(\ln \frac{1+N}{1+\text{df}_t}\right)$$

8) Robustness and Sensitivity Analysis

Four robustness checks were carried out. Firstly, all inferential tests were re-run after excluding the three incidents with disputed record counts. Secondly, the risk index was recomputed under two alternative weighting schemes, an unweighted scheme (1.0, 1.0, 1.0) and a steeper scheme (1.0, 2.0, 3.0), and the resulting rankings were compared with the base ranking using Spearman's rank correlation. Thirdly, the crypto versus non-crypto loss difference was checked with a bootstrap confidence interval, which does not rely on distributional assumptions. Fourthly, the annual incident count was tested for a monotonic trend to check whether the rise in recent incidents reflects the full period or only a few years.

9) Software and Reproducibility

All analysis was performed in Python 3 using pandas for data handling, NumPy and SciPy for statistical computation and optimisation, and scikit-learn for TF-IDF weighting and Cohen's kappa. The logistic regression and the OLS model with HC3 errors were implemented directly from their matrix definitions, and all permutation and bootstrap procedures used a fixed random seed. The coded dataset and scripts are available as described in the Declarations.

4. Results

1) Descriptive Overview

The dataset contains 47 incidents, of which 31 report records exposed and 27 report a financial loss. North America accounts for 23 incidents (48.9%), Asia-Pacific for 13 (27.7%), Europe for 5, and Africa & Middle East and Global/Offshore platforms for 3 each. The number of documented incidents rises across the four periods (4 in 2008–2015, 12 in 2016–2019, 15 in 2020–2022 and 16 in 2023–2025), which partly reflects the growth of the FinTech sector and partly the stronger disclosure rules of recent years. Table 4 summarises the incidents by sub-sector.

Table 4: Incidents and impact by FinTech sub-sector

Sub-sector	Incidents (%)	Median records (millions)	Median loss (US\$ millions)	Total reported loss (US\$ millions)
Crypto & DeFi	13 (27.7%)	0.07 (1 case)	290.0	4,956.0
Banks & credit unions	11 (23.4%)	7.6	52.5	388.5
Payments & neobanks	10 (21.3%)	7.5	15.0	253.3
Brokerage & insurance	7 (14.9%)	2.6	1.5	1.5
Lending, mortgage & credit	6 (12.8%)	20.4	38.5	802.0

Two very different kinds of harm can be clearly seen in Table 4. Crypto and DeFi incidents rarely expose customer records but cause very large direct theft of funds, while lenders, credit bureaus, banks and payment firms mainly suffer large exposures of personal and financial data. Crypto and DeFi

platforms account for 77.4% of all reported financial loss in the dataset.

2) Breach Patterns over Time (RQ1)

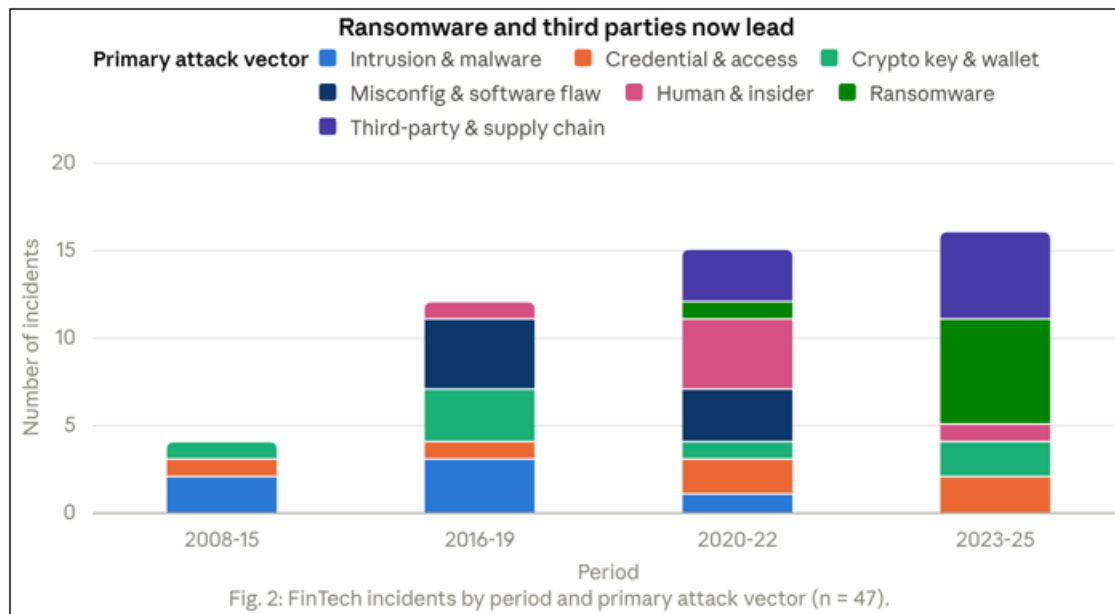


Fig. 2: FinTech incidents by period and primary attack vector (n = 47).

Author's FinTech incident dataset · 47 incidents · 2008–2025

Fig. 2 shows a clear change in how FinTech firms are being breached. Intrusion and malware attacks, such as the Heartland, Bangladesh Bank and Cosmos Bank cases, dominated the early periods but do not appear as a primary vector in 2023–2025. Misconfiguration and software flaws

peaked in 2016–2019 (Equifax, Capital One, First American). In contrast, ransomware rose from zero incidents before 2020 to six of the sixteen incidents (37.5%) in 2023–2025, and third-party and supply-chain compromise rose from zero to five.

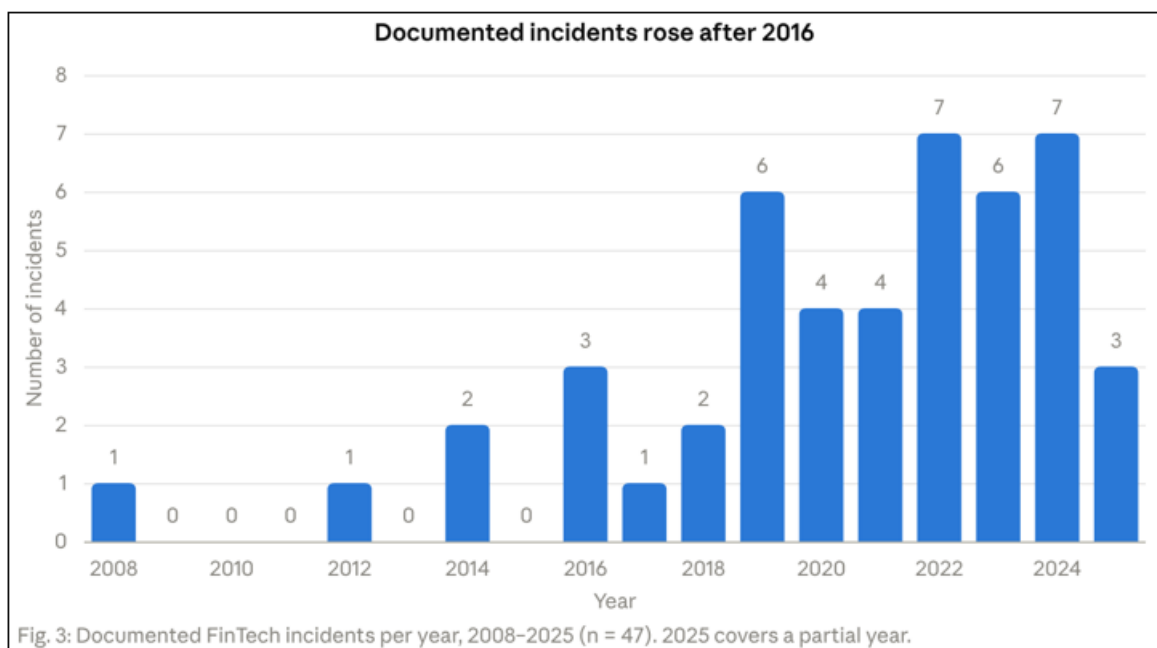


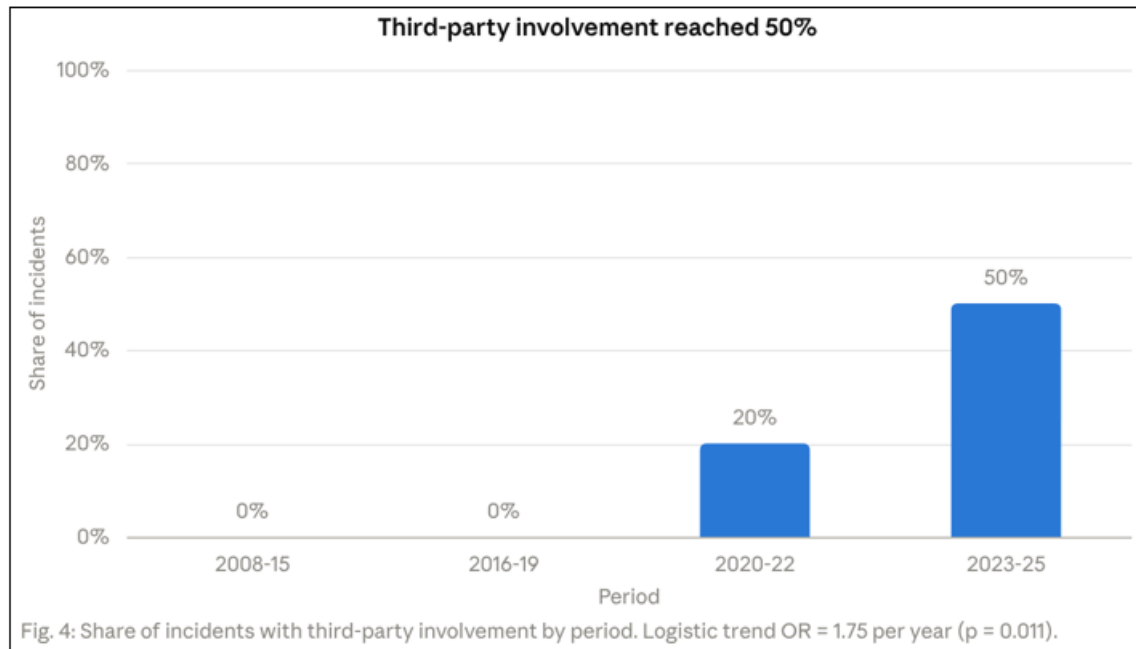
Fig. 3: Documented FinTech incidents per year, 2008–2025 (n = 47). 2025 covers a partial year.

Author's FinTech incident dataset · annual counts · 2008–2025

The annual number of documented incidents shows a strong upward trend (Spearman $\rho = 0.83$, $p < 0.001$), as shown in Fig. 3. Only four incidents were recorded between 2008 and 2015, compared with 31 between 2020 and 2025. The lower count for 2025 reflects the fact that the year was only partially observed at the time of data collection and should not be read as a decline.

When third-party involvement is measured as a root-cause factor, which also captures cases where a vendor was part of

the attack chain, its share rises from 0% in 2008–2015 and 2016–2019 to 20% in 2020–2022 and 50% in 2023–2025. A logistic regression confirms this as a significant trend: each additional year increases the odds of third-party involvement by about 75% (OR = 1.75, 95% CI 1.14–2.71, $p = 0.011$). This result is consistent with the doubling of third-party involvement reported by Verizon for 2025 [15]. The share of incidents involving a human factor stayed between 25% and 33% from 2016 onwards.



Author's FinTech incident dataset · third-party share by period

The result is robust to the exclusion of the three disputed incidents (OR = 1.71, 95% CI 1.09–2.70, $p = 0.020$). Fig. 4 shows that the increase is concentrated in the most recent period, which coincides with the MOVEit, Snowflake and outsourced-support compromises that affected several financial firms at once.

Breach size, however, did not grow over time. The Spearman correlation between year and log records is negative ($\rho = -0.36$, $p = 0.047$), meaning that recent incidents in the dataset

expose somewhat fewer records than the mega-breaches of 2008–2019, and there is no significant trend in log financial loss ($\rho = -0.01$, $p = 0.96$). This supports the finding of Edwards et al. [8] that breach size does not necessarily grow over time, and it suggests that the risk in FinTech is changing in its route of attack rather than in the size of individual events.

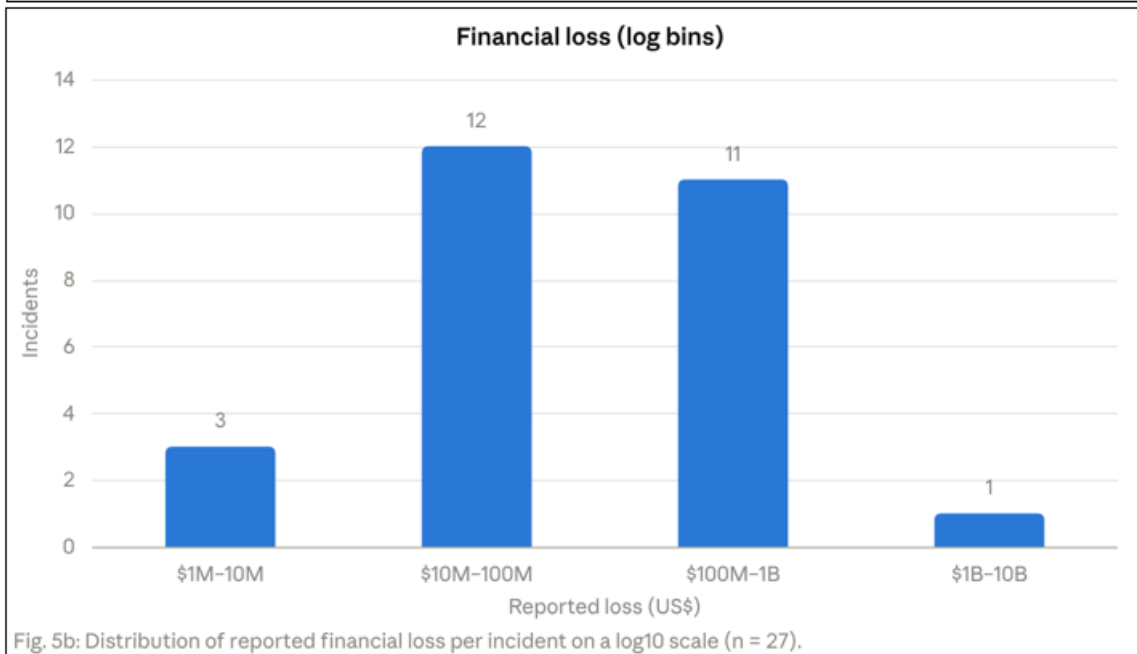
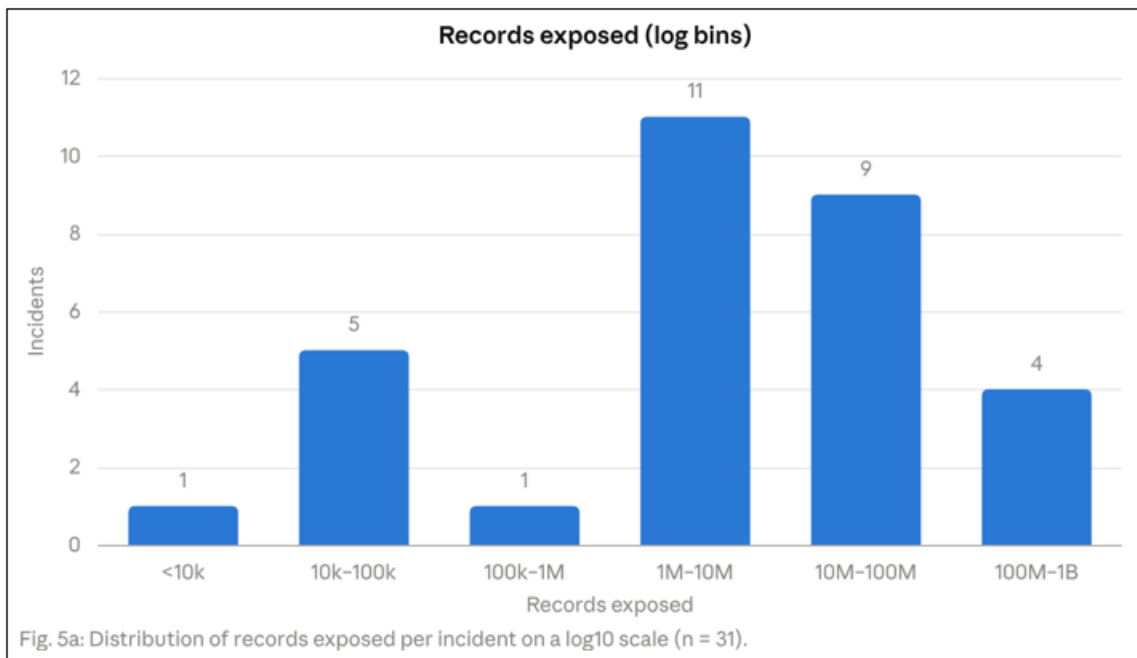
3) Distribution of Breach Impact (RQ3)

Table 5: Distribution of breach impact

Measure	Records exposed (millions)	Financial loss (US\$ millions)
Incidents with data	31	27
Total	1,691.1	6,401.3
Mean	54.6	237.1
Median	7.6	81.0
Interquartile range	1.25–32.5	26.0–305.0
Maximum	885.0 (First American)	1,500.0 (Bybit)
Skewness	4.76	2.27
Share held by top 10% (3 incidents)	68.7%	44.1%
Shapiro–Wilk p (raw / log10)	< 0.001 / 0.19	< 0.001 / 0.19

Both impact measures are extremely heavy-tailed. The mean number of records exposed is more than seven times the median, and just three incidents (First American, Heartland and Equifax) hold 68.7% of all exposed records. The Shapiro–Wilk test strongly rejects normality for the raw values but not for the log-transformed values ($p = 0.19$ in both

cases), which indicates that FinTech breach impact follows an approximately log-normal distribution, similar to the pattern found in cross-industry breach data [8], [9]. This has an important practical meaning: average-based cost estimates, such as a single industry average, can seriously understate the tail risk that a FinTech firm faces.



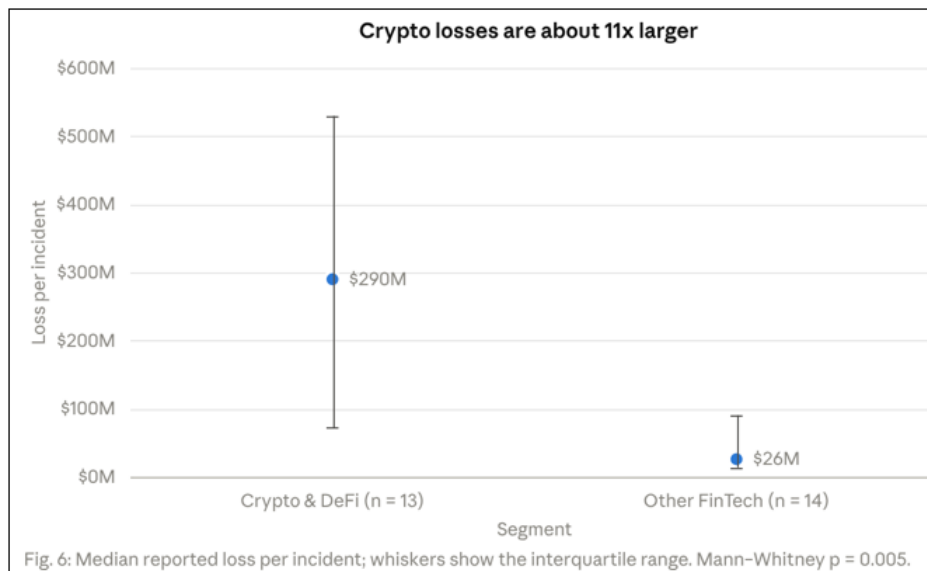
Author's FinTech incident dataset · impact distributions

Fig. 5 presents both distributions on logarithmic bins. Records exposed cluster between 1 million and 100 million per incident, with a thin tail reaching nearly one billion, while financial losses cluster between US\$10 million and US\$1 billion. The roughly symmetric shape of both distributions on the log scale is consistent with the Shapiro–Wilk results in Table 5.

4) Attack Vectors, Sub-sectors and Impact (RQ2, RQ3)

The distribution of attack vectors differs significantly across sub-sectors (permutation $\chi^2 = 39.08$, $p = 0.020$, Cramér's $V =$

0.46), which is a strong association. The clearest example is that all seven cryptographic key and wallet compromises occurred in crypto and DeFi platforms and none occurred elsewhere (Fisher's exact test, $p < 0.001$). Crypto and DeFi incidents also caused significantly higher financial losses, with a median of US\$290 million compared to US\$26 million for other FinTech segments (Mann–Whitney $U = 149$, $p = 0.005$).



Author's FinTech incident dataset · 27 incidents with reported loss

The difference in Fig. 6 is also supported by the bootstrap analysis, which gives a 95% confidence interval of US\$23.5 million to US\$509 million for the difference in median loss, so the interval clearly excludes zero. The interquartile ranges of the two groups barely overlap.

The number of records exposed did not differ significantly across attack vectors (Kruskal–Wallis $H = 4.46$, $p = 0.49$),

although the medians vary widely: misconfiguration and software flaw incidents had the highest median exposure (126.5 million records, $n = 4$), followed by credential and access compromise (35.0 million, $n = 5$), while intrusion and malware incidents had a median of 1.5 million ($n = 4$). Table 6 compares incidents with and without each root-cause factor.

Table 6: Records exposed by root-cause factor (Mann–Whitney U tests)

Factor	Incidents with factor (n)	Median records with (millions)	Median records without (millions)	p-value	Rank-biserial r
Misconfiguration or software flaw	8	60.9	7.5	0.119	0.38
Weak authentication	9	14.0	7.2	0.695	0.10
Human factor	8	7.9	7.5	0.542	0.15
Third-party involvement	9	7.5	8.9	0.601	-0.13

None of the individual factors reached statistical significance, but misconfiguration and software flaws show a medium effect size ($r = 0.38$), with a median exposure about eight times larger than other incidents. This suggests that technical flaws in internet-facing systems, such as an unpatched web framework or a misconfigured cloud firewall, tend to expose entire databases at once, whereas human and third-party attacks tend to expose a narrower slice of data.

Table 7: OLS regression of log10 records exposed (HC3 robust standard errors)

Predictor	Coefficient	Robust SE	p-value
Year	-0.082	0.062	0.197
Human factor	0.034	0.492	0.946
Third-party involvement	0.057	0.527	0.915
Weak authentication	0.271	0.513	0.602
Misconfiguration or software flaw	0.368	0.702	0.604
$n = 31$, $R^2 = 0.10$, adjusted $R^2 = -0.08$, $F = 0.58$ ($p = 0.72$)			

The multivariable model in Table 7 explains very little of the variation in breach size ($R^2 = 0.10$), and removing the three disputed incidents does not change this conclusion ($R^2 = 0.16$, $n = 28$, no significant predictors). This is itself an important finding: the root cause of a FinTech breach is informative

about how it happened, but reveals very little about how large it will be. Breach size appears to be driven mainly by how much data the compromised system held, which supports data minimisation and segmentation as direct ways of limiting impact.

5) Text Analysis of Incident Narratives

The dictionary-based coding showed almost perfect agreement with manual coding [24], with a pooled Cohen's κ of 0.91 and 96.8% raw agreement across all 282 incident–factor decisions. Agreement by factor is shown in Table 8.

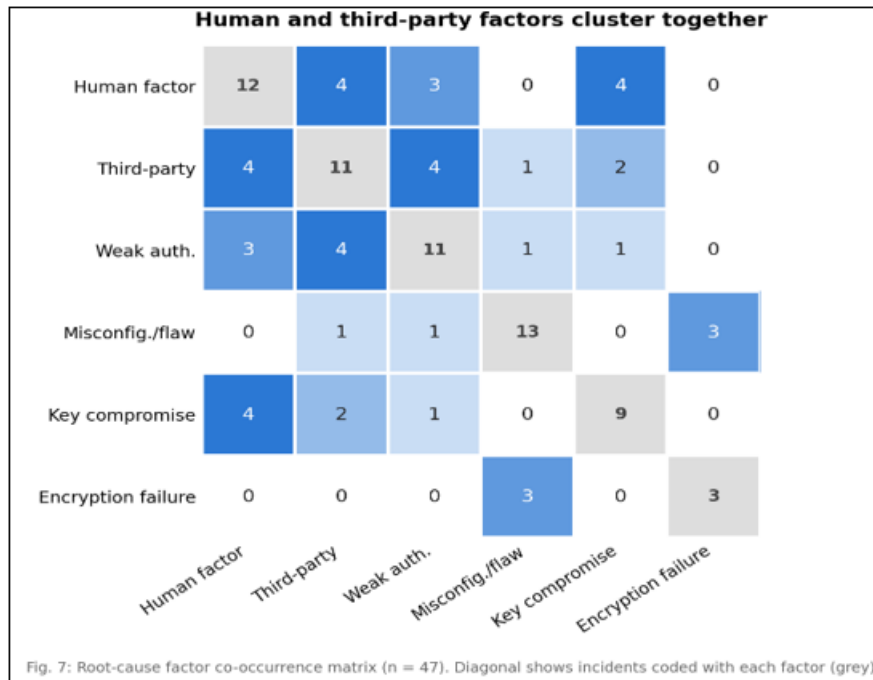
Table 8: Agreement between manual and dictionary-based coding

Root-cause factor	Incidents coded (manual)	Cohen's κ	Raw agreement
Misconfiguration or software flaw	13	0.95	97.9%
Human factor	12	0.94	97.9%
Third-party involvement	11	0.94	97.9%
Weak authentication	11	0.94	97.9%
Cryptographic key compromise	9	0.80	93.6%
Encryption failure	3	0.73	95.7%

The content analysis also shows that FinTech incidents are rarely caused by a single weakness. 17 of the 47 incidents (36.2%) involved two or more root-cause factors, with an

average of 1.26 factors per incident. The most frequent combinations were human factor with third-party involvement (4 incidents, for example Bybit and Coinbase), human factor with key compromise (4, for example Ronin Network and Coincheck), and third-party involvement with

weak authentication (4, for example Latitude Financial and Santander). All three encryption failures (Heartland, Equifax and Monzo) occurred together with a misconfiguration or software flaw.



Author's FinTech incident dataset · manual root-cause coding

The co-occurrence structure in Fig. 7 shows two broad clusters. The first links the human factor, third-party involvement, weak authentication and key compromise, reflecting attacks that exploit people and trust relationships. The second links misconfiguration and software flaws with encryption failure, reflecting purely technical weaknesses. The two clusters rarely overlap, since the human factor never co-occurs with misconfiguration in the dataset.

The TF-IDF analysis shows a clear shift in the vocabulary of incidents. For 2008–2019, the highest-weighted terms were "malware", "hot wallet", "web application", "SWIFT", "bitcoin" and "card", which describe direct attacks on a firm's own systems. For 2020–2025, the highest-weighted terms were "ransomware", "social engineering", "third party", "customer data" and "exfiltrated". Table 9 shows the share of narratives containing key threat terms in each era.

Table 9: Share of incident narratives containing key terms

Term group	2008–2019 (n = 16)	2020–2025 (n = 31)
Third-party, vendor, provider, outsourced	0%	29%
Ransomware	0%	23%
Smart contract, bridge	0%	13%
Credential, password, login	12%	16%
Cloud	6%	10%
Key	31%	13%
Malware	31%	0%

The text evidence therefore supports the quantitative trend: the language of FinTech breaches has moved away from malware on a firm's own infrastructure towards ransomware, extortion and compromise through partners and people.

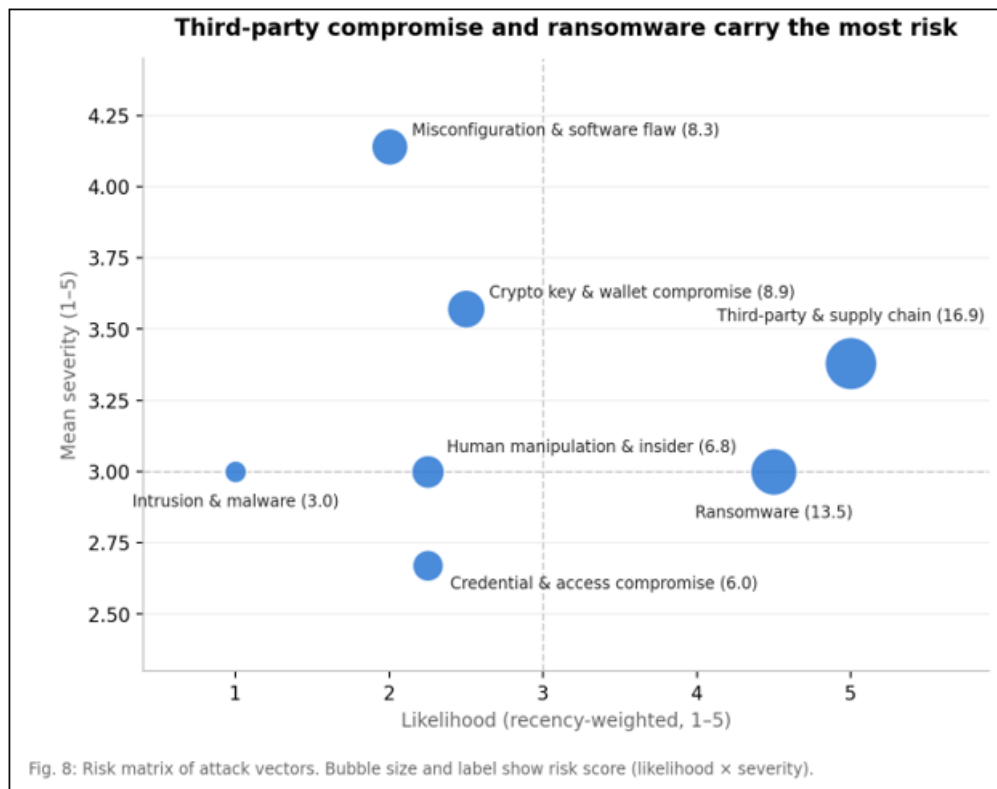
6) Risk Index (RQ4)

Table 10: Recency-weighted risk index by attack vector

Rank	Attack vector	Incidents	Likelihood (1–5)	Mean severity (1–5)	Risk score
1	Third-party & supply chain	8	5.00	3.38	16.88
2	Ransomware	7	4.50	3.00	13.50
3	Crypto key & wallet compromise	7	2.50	3.57	8.93
4	Misconfiguration & software flaw	7	2.00	4.14	8.29
5	Human manipulation & insider	6	2.25	3.00	6.75
6	Credential & access compromise	6	2.25	2.67	6.00
7	Intrusion & malware	6	1.00	3.00	3.00

Third-party and supply-chain compromise ranks first because it is both the most frequent vector in recent years and has above-average severity. Ransomware ranks second, mainly because of its high recent frequency. Misconfiguration and

software flaws have the highest average severity (4.14) but a lower recent likelihood, which makes them a "low-frequency, high-impact" risk that should not be ignored.



Author's FinTech incident dataset · risk index by attack vector

Fig. 8 places each attack vector on a likelihood–impact plane. Third-party compromise and ransomware occupy the high-likelihood region, while misconfiguration and software flaws

occupy the high-impact region. Intrusion and malware, which dominated the early literature, now sit in the lowest-risk corner.

Table 11: Recency-weighted risk index by sub-sector

Rank	Sub-sector	Incidents	Likelihood (1–5)	Mean severity (1–5)	Risk score
1	Crypto & DeFi	13	5.00	3.69	18.46
2	Banks & credit unions	11	3.44	3.00	10.33
3	Payments & neobanks	10	2.78	2.70	7.50
4	Brokerage & insurance	7	1.89	3.00	5.67
5	Lending, mortgage & credit	6	1.00	4.17	4.17

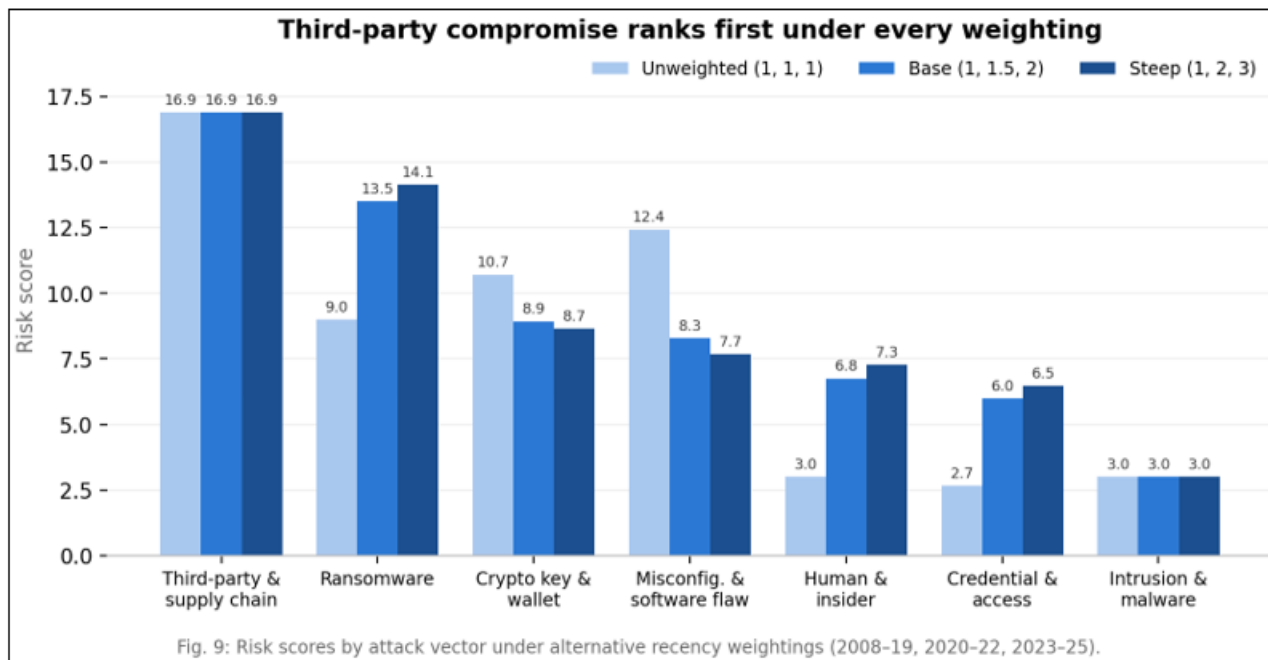
Crypto and DeFi platforms carry the highest overall risk score, combining the highest frequency with high severity. Lending, mortgage and credit firms have the highest mean severity (4.17), since credit bureaus and mortgage servicers hold very large volumes of sensitive personal data, but a lower frequency in the dataset.

7) Robustness and Sensitivity Analysis

The main results were tested against the robustness checks described in Section III-H. Table 12 summarises how each result changed.

Table 12: Summary of robustness checks

Check	Base result	Result under check	Conclusion
Exclude disputed incidents: third-party trend	OR = 1.75, $p = 0.011$	OR = 1.71, 95% CI 1.09–2.70, $p = 0.020$	Robust
Exclude disputed incidents: OLS model	$R^2 = 0.10$, no significant predictor	$R^2 = 0.16$, no significant predictor ($n = 28$)	Robust
Bootstrap: crypto vs other median loss	Mann–Whitney $p = 0.005$	95% CI for difference US\$23.5m–US\$509m	Robust
Annual count trend	Rise across four periods	Spearman $\rho = 0.83$, $p < 0.001$ over 18 years	Robust
Risk ranking: steep weights (1, 2, 3)	Base ranking	Identical ranking ($\rho = 1.00$)	Robust
Risk ranking: unweighted (1, 1, 1)	Base ranking	$\rho = 0.79$, $p = 0.033$; third-party still ranked first	Largely robust



Author's FinTech incident dataset · three weighting schemes

Fig. 9 shows the effect of the weighting scheme on the risk index. Third-party and supply-chain compromise ranks first under all three schemes. Without recency weights, misconfiguration and software flaws move up to second place and ransomware falls to fourth, because most misconfiguration incidents occurred before 2020 while most ransomware incidents occurred after it. This confirms that the

recency weights change the ordering in a way that reflects the recent shift in the threat landscape, but the top-ranked risk does not depend on them.

8) Summary of Hypothesis Tests

Table 13 summarises the outcome of each hypothesis.

Table 13: Summary of hypothesis tests

Hypothesis	Test	Result	Outcome
H1: Third-party involvement has increased over time	Logistic regression on year	OR = 1.75 per year, 95% CI 1.14–2.71, $p = 0.011$	Supported
H2: Breach size has increased over time	Spearman correlation, year vs log10 records	$\rho = -0.36$, $p = 0.047$ (negative direction)	Not supported
H3: Impact is heavy-tailed and log-normal	Shapiro–Wilk on raw and log10 values; top-10% share	Raw $p < 0.001$; log $p = 0.19$; top 3 incidents hold 68.7% of records	Supported
H4: Attack vectors differ across sub-sectors	Permutation chi-square (10,000 permutations)	$\chi^2 = 39.08$, $p = 0.020$, Cramér's $V = 0.46$	Supported
H5: Crypto and DeFi losses are larger	Mann–Whitney U	Median US\$290m vs US\$26m, $p = 0.005$	Supported
H6: Root-cause factors are associated with breach size	Mann–Whitney U; OLS with HC3 errors	No factor significant; $R^2 = 0.10$, F-test $p = 0.72$	Not supported

Four of the six hypotheses are supported. The rejection of H2 and H6 is equally informative, since it indicates that FinTech cyber risk is changing in its route of attack rather than in the size of individual breaches, and that breach size depends on the data held rather than on the root cause.

5. Discussion

1) From Perimeter Risk to Ecosystem Risk

The most important finding of this study is that FinTech cyber risk has moved from the perimeter of individual firms to the wider ecosystem that connects them. Before 2020, the incidents in the dataset were mostly direct attacks on a firm's own infrastructure, whether through malware on payment systems (Heartland, Bangladesh Bank, Cosmos Bank) or through flaws in its own web applications and cloud configuration (Equifax, Capital One). From 2020 onwards, half of the most recent incidents involved a third party, and

the attackers increasingly reached their targets through a vendor's file-transfer software (Genworth via MOVEit), an outsourced support team (Coinbase), a cloud data platform (Santander), a service provider's credentials (Latitude Financial) or a wallet provider's developer machine (Bybit). The Evolve Bank & Trust case shows the reverse direction as well: a breach at one banking-as-a-service provider exposed the customers of the FinTech firms that depended on it.

This pattern supports the argument of Kopp et al. [12] and the IMF [13] that concentration in shared service providers turns cyber risk into a system-wide problem, and it matches the doubling of third-party involvement in the Verizon DBIR [15] and the rise of supply-chain compromise to the second most common vector in the IBM study [14]. For FinTech firms, which depend on partner banks, cloud providers, KYC vendors and APIs by design, this means that a firm's security is only as strong as the weakest partner in its chain. It also explains why regulators have started to address third-party

risk directly, as seen in DORA [19] and in the RBI's IT governance directions [17].

2) Two Different Risk Profiles within FinTech

The results show that "FinTech" is not one uniform risk category. Traditional and digital financial firms (banks, payment firms, lenders and credit bureaus) mainly face data-exposure risk: breaches that leak millions of customer records, leading to identity theft, fraud, regulatory penalties and settlement costs. Crypto and DeFi platforms, on the other hand, face direct theft risk: a single compromised key or flawed smart contract can move hundreds of millions of dollars in minutes, and the funds are usually not recoverable, a pattern that blockchain analytics firms have also tracked across the crypto sector [16]. This difference was statistically significant (median loss of US\$290 million versus US\$26 million) and all key and wallet compromises occurred in the crypto segment. As a result, security priorities must be different: data minimisation, encryption and access control for the first group, and key management, multi-signature governance, transaction-signing integrity and smart contract auditing for the second.

3) The Human and Technical Factors Work Together

The content analysis shows that more than a third of incidents involved two or more root-cause factors, and that the most damaging combinations joined a human weakness to a technical one. In the Ronin Network and Bybit cases, social engineering of a single individual led to the compromise of signing keys or signing interfaces worth over US\$600 million and US\$1.5 billion respectively. In the Coinbase case, bribed outsourced support agents combined the human factor with third-party involvement. This supports the view in the literature [2], [3] and in Khullar [1] that employee awareness and technical controls cannot be treated separately. It also shows that multi-factor authentication, although important, is not enough on its own, since several recent attacks bypassed authentication entirely by targeting people who already had legitimate access.

4) What Drives the Size of a Breach

The regression and factor tests found that root causes explain very little of the variation in breach size, while the distribution of impact is heavy-tailed and log-normal. In practical terms, this means that the size of a breach depends mainly on how much sensitive data sits behind the compromised system, not on how the attacker got in. The largest data exposures (First American, Heartland, Equifax and Capital One) all involved a single flaw or misconfiguration that opened a very large central data store. For FinTech firms, this makes data minimisation, network segmentation and encryption at rest some of the most cost-effective controls, because they reduce the potential impact of every attack vector at once. This is consistent with the economic reasoning of Gordon and Loeb [6], which links the value of security investment to the potential loss from the information being protected.

5) Comparison with Prior Literature

Khullar [1], through a review of secondary literature, presented malware, phishing, data leakages and DDoS attacks as the main threats to FinTech and recommended countermeasures such as end-to-end encryption, MFA, AI-powered detection and blockchain. The present study both supports and refines those conclusions. It confirms that the human factor and encryption failures play a major role, but it shows through incident-level data that the threat landscape has since shifted towards ransomware and third-party compromise, which received less attention in earlier reviews [1], [3]. It also shows that the relative importance of threats differs across sub-sectors, which was not measured in previous FinTech-specific studies [2], [3].

6) Proposed FinTech Cyber Risk Assessment Framework

Based on these findings, this study proposes a five-stage FinTech Cyber Risk Assessment Framework (FCRAF) that FinTech firms, auditors and regulators can apply using their own incident and vendor data, as summarised in Table 14.

Table 14: The FinTech Cyber Risk Assessment Framework (FCRAF)

Stage	Purpose	Key activities	Evidence from this study
1. Profile	Identify the firm's risk type	Classify the firm as data-exposure dominant (banks, payments, lending) or theft dominant (crypto and DeFi); map where sensitive data and signing keys are held	Crypto losses 11× higher (median); all key compromises in crypto
2. Map the ecosystem	Make third-party risk visible	List all vendors, cloud platforms, BaaS partners and outsourced teams with access to data, systems or keys; rate each on access level and security controls	Third-party involvement rose to 50% of recent incidents (OR = 1.75 per year)
3. Score	Prioritise risks with data	Apply the recency-weighted likelihood × impact index to the firm's own and sector-wide incident data; rank attack vectors	Third-party (16.9) and ransomware (13.5) ranked highest
4. Limit impact	Reduce the tail	Data minimisation, segmentation, encryption at rest, withdrawal limits and multi-signature controls; model losses as heavy-tailed rather than using averages	Top 3 incidents hold 68.7% of exposed records; log-normal impact
5. Monitor and update	Keep the assessment current	Re-score quarterly, track new incident patterns through text analysis of disclosures, and test controls against multi-factor attack chains	Vocabulary shift towards ransomware and third parties; 36% of incidents multi-factor

The framework is designed to be practical and cumulative. Stages 1 and 2 can be completed by a small FinTech start-up with a simple spreadsheet, while Stages 3 to 5 allow larger firms and regulators to use statistical evidence to decide where limited security budgets should be spent first.

7) Theoretical and Practical Implications

Theoretical implications. Firstly, the study extends the finding of log-normal breach size [8], [9] from cross-industry data to the FinTech sector specifically, and shows that the same heavy-tailed pattern holds for direct financial theft as well as for records exposed. Secondly, the rising share of

third-party incidents provides incident-level evidence for the concentration argument of Kopp et al. [12], which had largely been made on conceptual grounds. Thirdly, the weak association between root cause and breach size suggests that models of cyber loss should separate the probability of compromise, which depends on the attack route, from the severity of compromise, which depends mainly on the volume of data or assets exposed.

Implications for FinTech firms. The results suggest three priorities. Vendor and partner security should be assessed with the same rigour as internal systems, including continuous monitoring of partner access and contractual rights to audit. Ransomware readiness, including offline backups, segmented networks and tested recovery plans, should be treated as a baseline requirement. Crypto and DeFi platforms should prioritise key management, multi-signature governance and independent verification of transaction-signing interfaces, since these controls address the costliest incidents in the dataset.

Implications for regulators and policymakers. The results support regulatory attention to critical third-party providers, as reflected in DORA [19], and to consistent and timely incident reporting. In India, CERT-In's 2022 directions require cyber incidents to be reported within six hours [36], and the RBI's IT governance directions [17] and the DPDP Act [18] strengthen accountability for regulated entities. Standardised public disclosure of attack vector, records affected and financial impact would allow studies of this kind to be repeated on larger and less biased datasets, which would in turn improve the evidence base for supervision.

6. Limitations and Future Scope

6.1 Limitations

This study has several limitations that should be kept in mind. Firstly, the dataset relies on publicly reported incidents, so it is affected by disclosure bias: large firms, US-based firms and firms under strict notification laws are over-represented (48.9% of incidents are from North America), and many smaller incidents, especially in developing markets, are never reported. Secondly, the sample of 47 incidents is modest, which limits statistical power, particularly for the regression model and for comparisons between individual attack vectors. Thirdly, impact figures are not perfectly comparable: records exposed and financial losses are reported inconsistently across firms, crypto losses are valued at the time of theft, and some record counts were claimed by attackers or disputed by companies, although a sensitivity analysis without the disputed cases did not change the conclusions. Fourthly, the incident narratives were compiled and manually coded by a single author, and the dictionary used for automated coding was developed with the codebook in view, so the high kappa value demonstrates consistency and reproducibility of the coding scheme rather than fully independent validation. Finally, the recency weights and severity bands used in the risk index involve judgement, and different choices could change the exact scores, although the top-ranked risks were stable across the time periods analysed.

6.2 Future Scope

Future research can extend this work in several ways. The dataset can be expanded using regulatory breach registers and cyber insurance loss data, with a particular focus on Indian and emerging-market FinTech firms, where disclosure is now improving under the Digital Personal Data Protection Act [18]. A second coder and a larger corpus of full-length breach notifications would allow more advanced text analysis, such as topic modelling and sentiment analysis of how firms communicate breaches. Extreme value theory could be used to model the tail of FinTech losses more precisely, following Eling and Wirfs [10]. Finally, the proposed framework can be tested in practice with FinTech start-ups to measure whether it improves their security decisions.

7. Conclusion

This study set out to move FinTech cybersecurity research from description towards measurement, by building a structured dataset of 47 FinTech incidents from 2008 to 2025 and analysing it with both statistical methods and text-based content analysis. The findings show that FinTech cyber risk has changed significantly over time. Direct malware and perimeter attacks, which dominated the early years, have given way to ransomware and to attacks that travel through third parties, with third-party involvement rising from none of the incidents before 2020 to half of those in 2023–2025. Breach impact is extremely heavy-tailed and log-normal, so a small number of events cause most of the damage, and the size of a breach depends more on how much data a compromised system holds than on how the attacker got in. Furthermore, the FinTech sector contains two distinct risk profiles: data-exposure risk for banks, payment firms and lenders, and direct theft risk for crypto and DeFi platforms, which suffer significantly larger financial losses per incident.

The recency-weighted risk index identifies third-party and supply-chain compromise and ransomware as the highest-priority attack vectors, and crypto and DeFi as the highest-risk sub-sector. Based on these results, this study proposes a five-stage FinTech Cyber Risk Assessment Framework built around profiling, ecosystem mapping, data-driven scoring, impact limitation and continuous monitoring. For FinTech firms, the key message is that security must extend beyond their own systems to every partner that can reach their data, systems or keys. For regulators, the results support stronger oversight of shared service providers and clearer, more consistent breach disclosure, since better data is the foundation of better risk assessment. Ultimately, as financial services become more digital and more connected all over the world, protecting consumer trust will depend on managing cybersecurity risk not only within individual firms but across the entire FinTech ecosystem.

Declarations

Data availability: The incident dataset, coding scheme and Python analysis scripts supporting the findings of this study are provided in Appendices A and B and are available from the author upon reasonable request.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests: The author declares no competing interests.

Ethics statement: This study uses only publicly available information on organisational security incidents and does not involve human participants or personal data about individuals.

References

- [1] Khullar A. Data Breaches and Cybersecurity in Fintech – A Comprehensive Analysis. *International Journal of Innovative Science and Research Technology*. 2025;10(4):1116–27. Available from: <https://doi.org/10.38124/ijisrt/25apr1107>
- [2] Javaheri D, Fahmideh M, Chizari H, Lalbakhsh P, Hur J. Cybersecurity threats in FinTech: A systematic review. *Expert Systems with Applications*. 2024; 241: 122697.
- [3] Olaiya OP, Adesoga TO, Ojo A, Olagunju OD, Ajayi OO, Adebayo YO. Cybersecurity strategies in fintech: Safeguarding financial data and assets. *GSC Advanced Research and Reviews*. 2024;20(1). Available from: <https://doi.org/10.30574/gscarr.2024.20.1.0241>
- [4] Kshetri N. Cybercrime and cybersecurity in India: Causes, consequences and implications for the future. *Crime, Law and Social Change*. 2016;66(3):313–38.
- [5] Najaf K, Mostafiz MI, Najaf R. Fintech firms and banks sustainability: Why cybersecurity risk matters? *International Journal of Financial Engineering*. 2021.
- [6] Gordon LA, Loeb MP. The economics of information security investment. *ACM Transactions on Information and System Security*. 2002;5(4):438–57.
- [7] Romanosky S. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*. 2016;2(2):121–35.
- [8] Edwards B, Hofmeyr S, Forrest S. Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*. 2016;2(1):3–14.
- [9] Wheatley S, Maillart T, Sornette D. The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B*. 2016; 89: 7.
- [10] Eling M, Wirfs J. What are the actual costs of cyber risk events? *European Journal of Operational Research*. 2019;272(3):1109–19.
- [11] Aldasoro I, Gambacorta L, Giudici P, Leach T. The drivers of cyber risk. *Journal of Financial Stability*. 2022; 60: 100989.
- [12] Kopp E, Kaffenberger L, Wilson C. Cyber Risk, Market Failures, and Financial Stability. IMF Working Paper WP/17/185. International Monetary Fund; 2017.
- [13] International Monetary Fund. Cyber risk: A growing concern for macrofinancial stability. In: *Global Financial Stability Report*, April 2024. Chapter 3. IMF; 2024.
- [14] IBM Security, Ponemon Institute. Cost of a Data Breach Report 2025. IBM; 2025. Figures as summarised by CSO Online and Northdoor.
- [15] Verizon Business. 2025 Data Breach Investigations Report. Verizon; 2025. Press release; Executive summary.
- [16] Chainalysis. The 2025 Crypto Crime Report. Chainalysis Inc.; 2025.
- [17] Reserve Bank of India. Master Direction on Information Technology Governance, Risk, Controls and Assurance Practices. RBI; 2023.
- [18] Government of India. The Digital Personal Data Protection Act, 2023 (No. 22 of 2023). 2023.
- [19] European Union. Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (DORA). *Official Journal of the European Union*. 2022; L 333.
- [20] National Institute of Standards and Technology. The NIST Cybersecurity Framework (CSF) 2.0. NIST CSWP 29. NIST; 2024.
- [21] National Institute of Standards and Technology. Guide for Conducting Risk Assessments. NIST Special Publication 800-30 Rev. 1. NIST; 2012.
- [22] Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qualitative Health Research*. 2005;15(9):1277–88.
- [23] Krippendorff K. Content Analysis: An Introduction to Its Methodology. 4th ed. SAGE Publications; 2018.
- [24] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–74.
- [25] Gosha P. Rise of Fintech and its Challenges: A Study on Cybersecurity Threats to Fintech Industry. Indian Institute of Foreign Trade, Delhi.
- [26] ENISA. ENISA Threat Landscape: Finance Sector. European Union Agency for Cybersecurity; 2025.
- [27] Privacy Rights Clearinghouse. Data Breach Chronology [Internet]. Available from: <https://privacyrights.org>
- [28] Financial Stability Board. Financial Stability Implications from FinTech: Supervisory and Regulatory Issues that Merit Authorities' Attention. FSB; 2017.
- [29] Arner DW, Barberis J, Buckley RP. The evolution of Fintech: A new post-crisis paradigm? *Georgetown Journal of International Law*. 2016;47(4):1271–319.
- [30] Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960;20(1):37–46.
- [31] Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*. 1988;24(5):513–23.
- [32] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*. 1952;47(260):583–621.
- [33] Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*. 1947;18(1):50–60.
- [34] Financial Stability Board, BIS Committee on the Global Financial System. FinTech Credit: Market Structure, Business Models and Financial Stability Implications. FSB; 2017.
- [35] Khullar A. A Computational Text Analysis of Stakeholder Framing in the Swachh Bharat Mission. Unpublished manuscript; 2026.
- [36] Indian Computer Emergency Response Team (CERT-In). Directions under sub-section (6) of section 70B of the Information Technology Act, 2000 relating to information security practices, procedure, prevention, response and reporting of cyber incidents. Ministry of

Electronics and Information Technology, Government
of India; 2022.

Appendix A: FinTech Incident Dataset (n = 47)

Factor codes: H = human factor; TP = third-party involvement; WA = weak authentication or credential control; MS = misconfiguration or software flaw; KC = cryptographic key compromise; EF = encryption failure. * = record count claimed by attackers or disputed by the company. Losses are funds stolen, ransom paid, settlements or regulatory penalties as publicly reported, in US\$ at the time; for First American only regulatory penalties are counted, and for Coinbase the figure is the midpoint of the company's own estimated cost range. "-" = not publicly reported.

ID	Year	Organisation	Country	Sub-sector	Primary vector	Records (m)	Loss (US\$ m)	Factors
F01	2008	Heartland Payment Systems	USA	Payments & neobanks	Intrusion & malware	130	140	MS, EF
F02	2012	Global Payments	USA	Payments & neobanks	Intrusion & malware	1.5	94	-
F03	2014	JPMorgan Chase	USA	Banks & credit unions	Credential & access compromise	83	-	WA
F04	2014	Mt. Gox	Japan	Crypto & DeFi	Crypto key & wallet compromise	-	450	KC
F05	2016	Bangladesh Bank	Bangladesh	Banks & credit unions	Intrusion & malware	-	81	WA
F06	2016	Bitfinex	Hong Kong	Crypto & DeFi	Crypto key & wallet compromise	-	72	KC
F07	2016	Tesco Bank	UK	Banks & credit unions	Intrusion & malware	0.009	24	MS
F08	2017	Equifax	USA	Lending, mortgage & credit	Misconfiguration & software flaw	147	700	MS, EF
F09	2018	Coincheck	Japan	Crypto & DeFi	Crypto key & wallet compromise	-	530	H, KC
F10	2018	Cosmos Bank	India	Banks & credit unions	Intrusion & malware	-	13.5	-
F11	2019	Capital One	USA	Banks & credit unions	Misconfiguration & software flaw	106	270	MS
F12	2019	First American Financial	USA	Brokerage & insurance	Misconfiguration & software flaw	885	1.5	MS
F13	2019	Desjardins Group	Canada	Banks & credit unions	Human manipulation & insider	9.7	-	H
F14	2019	Binance	Global	Crypto & DeFi	Credential & access compromise	-	40	H, WA, KC
F15	2019	Monzo	UK	Payments & neobanks	Misconfiguration & software flaw	0.48	-	MS, EF
F16	2019	Upbit	South Korea	Crypto & DeFi	Crypto key & wallet compromise	-	49	KC
F17	2020	Travelex	UK	Payments & neobanks	Ransomware	-	2.3	MS
F18	2020	Juspay	India	Payments & neobanks	Credential & access compromise	35	-	WA, MS
F19	2020	Dave	USA	Payments & neobanks	Third-party & supply chain	7.5	-	TP
F20	2020	Experian South Africa	South Africa	Lending, mortgage & credit	Human manipulation & insider	24	-	H
F21	2021	MobiKwik *	India	Payments & neobanks	Credential & access compromise	99	-	-
F22	2021	Robinhood	USA	Brokerage & insurance	Human manipulation & insider	7	-	H
F23	2021	Upstox	India	Brokerage & insurance	Third-party & supply chain	2.5	-	TP, WA
F24	2021	Poly Network	Global	Crypto & DeFi	Misconfiguration & software flaw	-	611	MS
F25	2022	Revolut	Lithuania	Payments & neobanks	Human manipulation & insider	0.05	-	H
F26	2022	Cash App (Block)	USA	Payments & neobanks	Human manipulation & insider	8.2	15	H, WA
F27	2022	Ronin Network	Vietnam	Crypto & DeFi	Crypto key & wallet compromise	-	625	H, KC
F28	2022	Wormhole	Global	Crypto & DeFi	Misconfiguration & software flaw	-	320	MS
F29	2022	Nomad	USA	Crypto & DeFi	Misconfiguration & software flaw	-	190	MS
F30	2022	Flagstar Bank	USA	Banks & credit unions	Intrusion & malware	1.5	-	-
F31	2022	TransUnion South Africa *	South Africa	Lending, mortgage & credit	Third-party & supply chain	54	-	TP, WA
F32	2023	PayPal	USA	Payments & neobanks	Credential & access compromise	0.035	2	WA
F33	2023	Latitude Financial	Australia	Lending, mortgage & credit	Third-party & supply chain	14	50	H, TP, WA
F34	2023	Mr. Cooper	USA	Lending, mortgage & credit	Ransomware	14.7	25	-
F35	2023	Genworth Financial	USA	Brokerage & insurance	Third-party & supply chain	2.7	-	TP, MS

ID	Year	Organisation	Country	Sub-sector	Primary vector	Records (m)	Loss (US\$ m)	Factors
F36	2023	Bank of America (via Infosys McCamish)	USA	Banks & credit unions	Third-party & supply chain	0.057	-	TP
F37	2023	ICBC Financial Services	USA	Brokerage & insurance	Ransomware	-	-	MS
F38	2024	loanDepot	USA	Lending, mortgage & credit	Ransomware	16.9	27	-
F39	2024	Evolve Bank & Trust	USA	Banks & credit unions	Ransomware	7.6	-	H, TP
F40	2024	Patelco Credit Union	USA	Banks & credit unions	Ransomware	1	-	-
F41	2024	Prudential Financial	USA	Brokerage & insurance	Ransomware	2.5	-	-
F42	2024	Fidelity Investments	USA	Brokerage & insurance	Credential & access compromise	0.077	-	WA
F43	2024	Santander *	Spain	Banks & credit unions	Third-party & supply chain	30	-	TP, WA
F44	2024	WazirX	India	Crypto & DeFi	Crypto key & wallet compromise	-	235	TP, KC
F45	2025	Bybit	UAE	Crypto & DeFi	Third-party & supply chain	-	1,500	H, TP, KC
F46	2025	Coinbase	USA	Crypto & DeFi	Human manipulation & insider	0.069	290	H, TP
F47	2025	CoinDCX	India	Crypto & DeFi	Crypto key & wallet compromise	-	44	KC

Appendix B: Dictionary Coding Lexicon

The automated content analysis in Section III-D classified each incident narrative using the case-insensitive keyword patterns below. A factor was coded as present when at least one pattern, matched as a regular-expression substring, appeared in the narrative.

Factor	Code	Keyword patterns
Human factor	H	phish, social engineering, posing, bribed, employee clicked, phoned, malicious employee, former employee, fake job, deceived
Third-party involvement	TP	third-party, vendor, service provider, provider, partner fintech, outsourced, client
Weak authentication or credential control	WA	credential, login, password, two-factor, multi-factor, api keys, access key, cloud keys, account level access, access was not revoked
Misconfiguration or software flaw	MS	misconfigur, unpatched, vulnerabilit, flaw, coding error, sql injection, zero-day, faulty, weakness, upgrade
Cryptographic key compromise	KC	private key, wallet, keys, keeper, validator
Encryption failure	EF	unencrypted, encrypt, expired certificate, log files