Gallagher Re

TIDE Analysis

Leaked information could be music to the ears of cyber underwriters

Executive Summary

Technographic scanning offers insurers a powerful lens into the evolving landscape of commercial cyber risk. However, transforming raw technographic data into actionable underwriting insights remains a significant challenge. To address this, Gallagher Re developed TIDE: Technographic Insight Discovery Engine, a tool designed to pinpoint the most critical technographic metrics and guide insurers on their practical application.

This paper delves into the predictive power of leaked information as a key indicator of cyber claim frequency. We compare scores from Orpheus Cyber against technographic metrics and traditional firmographic factors, such as industry and revenue, to assess their relative impact on risk assessment.

Key Findings

- Leaked information on the surface, deep and dark web is a predictive driver of cyber claims, providing unique value beyond traditional firmographic data.
- In addition, we continue to see predictive power from technographic measures of the size of organisations' cyber footprints.
- Quality threat intelligence can add value when used to augment and interpret raw data sources.

How Insurers Benefit

- **Enhanced Risk Segmentation:** Identify and prioritise risks with the highest expected claim frequency to optimise portfolio management.
- **Smarter Underwriting:** Balance the proportion of insureds against their claim likelihood, enabling competitive pricing while minimising risk exposure or adjusting terms as needed.

Background

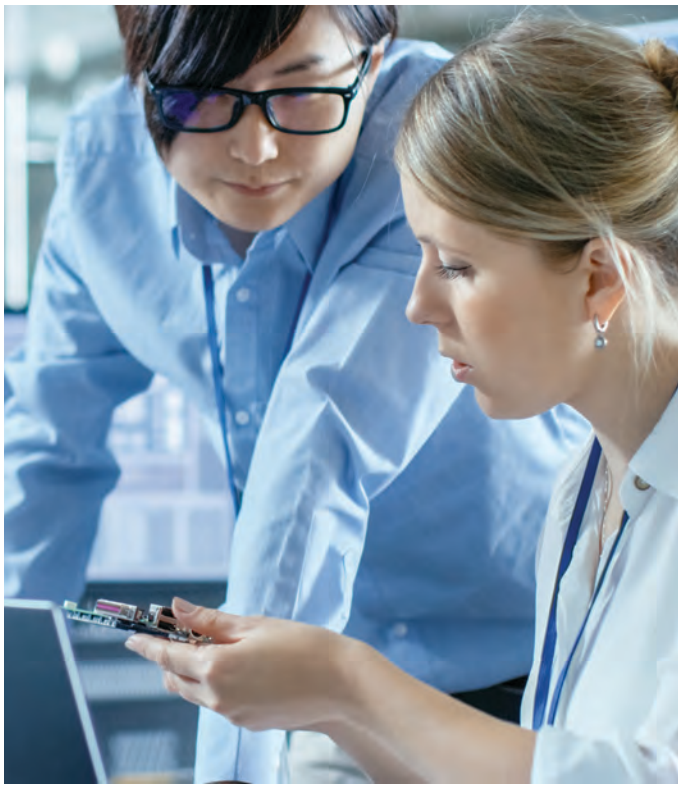
Firmographic features like Revenue, Country, and Industry have historically been central to cyber insurance pricing and risk selection; but in recent years insurers have increasingly looked to technographic data to provide additional insight and competitive advantage. The reality of using technographic data in practice is not straightforward: the breadth of this data is vast, its relationship with claims has not always been verified, and translating technographic data into actionable insight in a quantitative way is a challenge.

As a trusted reinsurance broker, Gallagher Re have a vested interest in providing our insurer clients with insights and analyses to derive the greatest value from technographic data:

- In 2022, "Looking from the Outside-In" outlined the various ways in which insurers were beginning to deploy these technologies and compared vendor factors and ability to capture individual cyber events.
- June 2023's "Can scanning technologies predict claims?" revealed the first results of applying machine learning techniques to understand the predictiveness of technographic data, highlighting the value of leveraging more granular technographic features.
- Last October, "Scanning the Horizon: How broadening our use of cybersecurity data can help insurers" built on our previous research to highlight changes and nuances to the factors which can signal cyber claims risk.

As a reinsurance broker, we are a trusted advisor to our clients which allows us to aggregate data anonymously, giving a wide visibility into firmographic, policy and claims data across the cyber (re)insurance market for their mutual benefit. We also partner with providers of technographic data to better understand the value of their data. Together, this positions us to provide a collaborative and complementary service to both (re) insurers and data vendors.

As a result of our research in this area, we have integrated external scanning data with firmographic and claims data to develop our portfolio quality and benchmarking product, TIDE, as well as enhancing the underwriting, monitoring, exposure management and pricing support services we provide to insurers.



What Is Leaked Information?

Leaked information refers to sensitive data, often stolen through data breaches, that is subsequently sold or shared on the internet. Additionally, this can also include information made publicly available such as information found on public websites, social media or third-party platforms. This sort of information typically includes login credentials, financial details and personal information, posing a significant risk to individuals and organisations as it can be used to gain access into privileged accounts, conduct fraud, and launch more sophisticated phishing campaigns.

Some of this leaked information is posted on the "surface web" (the part of the internet that is visible to search engines and normal web browsers) on easily accessible websites designed for hosting and sharing text and files, often known as "pastebins".

Other leaked information can be shared on the "deep web": parts of the internet (such as forums, messaging platforms or other restricted sites) which require an account, password or other credentials to access.

Leaked information is often posted on the "dark web": a part of the internet that is not indexed by regular search engines and requires special software, or browsers like Tor to access. It provides a high level of anonymity which has led to it being used for illegal activities like drug trafficking and the sale of stolen data. However, the dark web also has legitimate content and communities where it is used for privacy, activism, and journalism.

For this study, Gallagher Re analysed data from Orpheus Cyber. Orpheus differentiates itself in the cyber risk rating sector with its threat-led approach that integrates predictive machine learning with real-time threat intelligence from a wide range of sources. Importantly, this involves the combination of automated and 'human in the loop' processes to improve accuracy in correctly mapping any companies' attack surface and evaluate an organisation from a threat actors' viewpoint, The approach also includes positive weightings based on identification of correct use of effective technologies (such as quality VPN providers) and heavier emphasis on vulnerabilities and organisations which are actively being targeted by threat actors.

Our Approach

Gallagher Re's in-house Cyber Analytics team integrates large volumes of claims data with firmographic information in an extensive cyber insurance industry exposure database. We then combine this firmographic data and claims data with technographic data provided by cyber risk monitoring firms. For this study Orpheus supplied security scores for 47,000 companies associated with more than 1,000 claims.

Machine Learning (ML) is a useful tool to gain insights into how claim frequency correlates with technographic and firmographic data. Machine Learning algorithms can recognise and model patterns in data to make predictions about the future. The most performant algorithms generate models which are complex and difficult to explain: "black box models", but by applying additional Explainable AI (XAI) techniques we can "lift the lid" on the black box and gain insights into the patterns and correlations learnt by the models. These technologies enable our data scientists to identify which features are most important, and how the precise value of each feature influences the models' predictions (dependency).

Feature importance and dependency helps underwriters understand which technographic measures should be of primary concern and provide insights into how the specific value of a given technographic measure might influence risk quality. Our data scientists can train models on various combinations of technographic and firmographic data and evaluate the relative performance of each model. By comparing the performance of models trained on different datasets we can evaluate the additional marginal performance gain of using technographic data in combination with traditional firmographic data.

Our data science team trains thousands of models each year, studying combinations of claim type and feature data, and ensuring that claims are updated annually to capture the latest evolving threats and emerging trends.

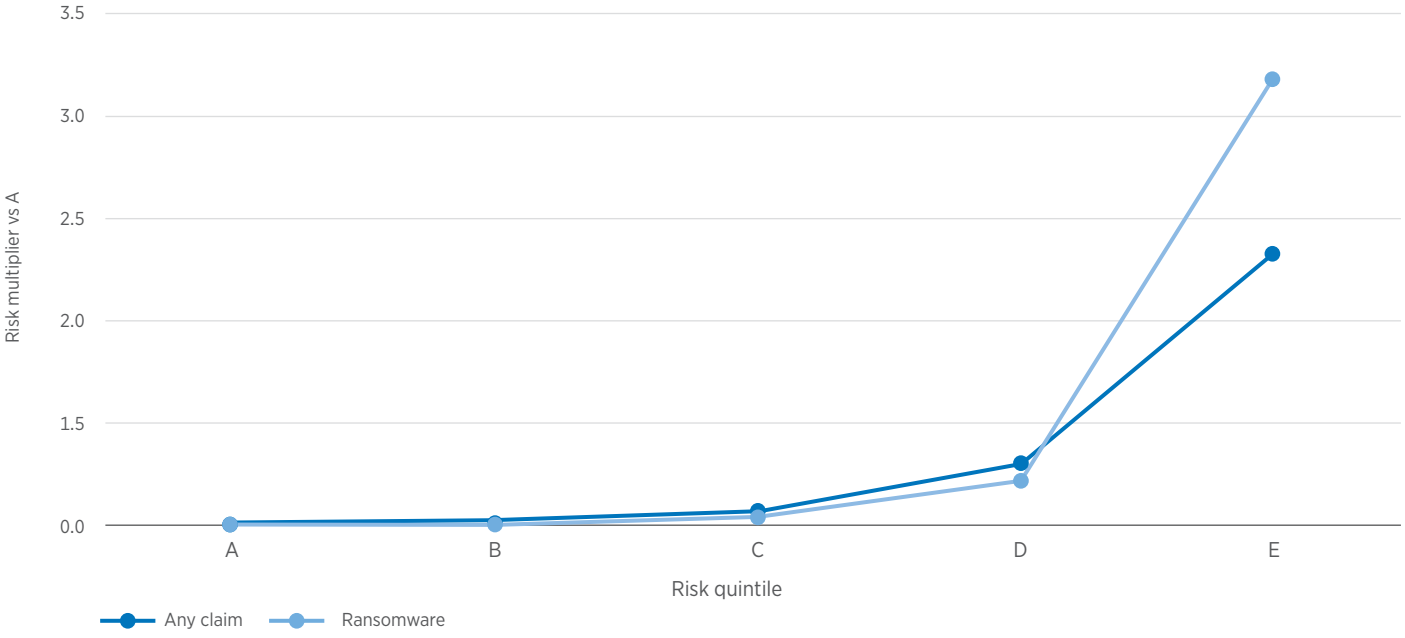
A critical tenet of our approach at Gallagher Re is to blend cyber security, data science, and insurance expertise. We understand the correlation is not causation and we view the findings of analysis through a critical lens to present a fair and pragmatic view.

Impact of Using Technographic Data

As in previous studies, we have found that technographic data excels at finding the greatest risks but sees little differentiation between the very best risks and those with a good-but-not-exceptional posture. This could be due to the nature of the security signals that are observed externally: on most meaningful security metrics, only a small proportion of organisations will display deficiencies, and it is only these highest-risk organisations that are flagged as being at a severely increased risk of claim.

This trend continues when we observe the impact of using Orpheus data to predict expected claim frequency. We grouped organisations into quintile buckets of risk as predicted by our model, from the lowest-risk 20% of organisations in quintile "A", to the highest-risk in quintile "E". We then compared average expected claim frequency per-quintile against the average of quintile "A" to demonstrate relative claim frequency risk of each quintile.

Figure 1: Relative claims frequencies for risk quintiles vs respective quintile A



This analysis again revealed the "hockey stick" comparison of risk likelihoods: those in quintiles A-C are predicted to see similar claims likelihoods (at most a 7% increase); those in quintile D are assessed at a marginally higher risk; but those in quintile E have a predicted risk likelihood which is multiple times higher than the best 20%. Using technographic data alone, these greatest 20% of risks are 2.3x more likely to suffer a claim than the best 20%.

Interestingly, there appears to be greater relative difference in the predicted ransomware claims frequency between the lowest and highest risk scores, the highest-risk 20% are 3.2x more likely to suffer a ransomware claim than those in quintile A. However, this could also be attributed to the lower volume of claims which are tagged as such by insurers, making this a less reliable conclusion. Improving the quality and detail of claims tagging is an ongoing priority for insurers to allow for deeper and more reliable insights on drivers of claims.

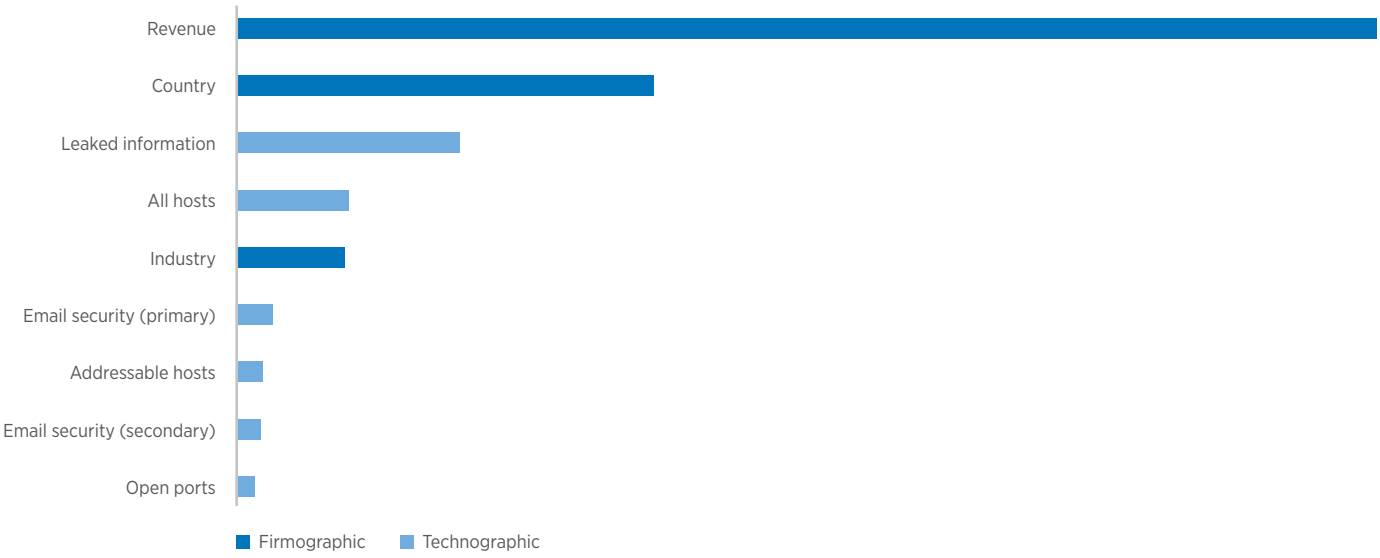
Insurers can use this data to identify insureds and applicants who are at a higher risk of a claim; and select, price, or set terms for these risks to ensure that they are not a cause for portfolio underperformance. Additionally, this allows insurers to be more competitive for the 60-80% of insureds who are not at this significantly higher risk of a claim.

Predictive and Additive Power of Leaked Information

Our model not only allows us to predict the expected claims frequency of a given risk based on all of the firmographic and technographic data available, but it also allows us to identify which specific features are most powerful in predicting claims.

Our previous analysis indicates that many risk features have predictive signal when examined in isolation, however we have previously found that they are often highly correlated with other features, or deliver minimal marginal predictive value vs. firmographic data alone.

Figure 2: Importance of features in predicting cyber claims



Although key firmographic features – Revenue, Country, and Industry – continue to be leading factors in predicting claims likelihood, specific technographic insights from Orpheus provide significant additive predictive power. Principal among these is Leaked Information, which is consistently associated with increased likelihood of suffering a claim.

Technographic measures of attack surface like Orpheus' All Hosts count continue to demonstrate value in an insurance context despite being relatively underserved by data vendors compared with their greater focus on vulnerability measures. Gallagher Re will publish a future publication exploring the valuable role that attack surface metrics can play in insurance underwriting.

The reasons why Leaked Information predicts a cyber claim are intuitive and well-known by threat intelligence professionals. Breached credentials can be used to gain access to user and privileged accounts in order to deploy attacks; higher volumes of breached information could suggest underlying security vulnerabilities that are allowing this information to be leaked; and higher levels of activity suggest that attackers are more interested in attacking the organisation. Additionally, the presence of exposed cloud storage, leaked source code, access keys or inadvertently shared sensitive files can further indicate lapses in secure configuration or operational discipline.

However, in Gallagher Re's experience, indicators of leaked information – whether breached credentials, data for sale, or hacker chatter on the deep and dark web – can often be "noisy", producing many results for most organisations. Posted credentials could be no longer valid or even reposted from a previous breach, data volumes and mentions are often evidence of a breach that has recently happened (as opposed to an impending one), and mentions can be highly correlated with company size and location.

Therefore, it was particularly interesting to see the predictive power of leaked information in the Orpheus dataset. Here, it is likely that the targeted and filtered interpretation of this data allows its security insight to translate into predictive power. For insurers, this highlights the difference that quality threat intelligence can add when used to augment and interpret raw data sources.

Oliver Church — CEO, Orpheus Security

Orpheus uses threat intelligence to highlight the most important issues in any companies' attack surface, weighting our scoring based on high-risk threat actor groups (such as Ransomware gangs' current known Techniques, Tactics and Procedures) to understand the likelihood of a compromise, and constantly adjust the prioritisation of different attack surface features in score weightings based on ever-changing attacker methodologies.

Threat intelligence is also used for evaluating the threat landscape surrounding an organisation. Orpheus' scoring considers factors such as a companies' industry sector, geographical presence and technology stack as well as the organisation's visible attack surface. This multifaceted analysis ensures that risk ratings reflect not only technical weaknesses but also the likelihood of targeted attacks.

Orpheus combines this threat-led approach with advanced machine learning and a commitment to deliver a range of predictive capabilities; from identifying in real time whether a CVE is currently being exploited or likely to be in the future (https://orpheus-cyber.com/wp-content/uploads/2025/01/Orpheus_Ministry-of-Defence_Case-Study.pdf), to uncovering emerging threats and anticipating shifts in attacker behaviour. This enables organisations to make proactive, intelligence-driven security decisions—to focus on what truly matters.

Calibrating Risk Selection and Pricing with Technographic Data

Gallagher Re's analysis also allows us to identify the points at which technographic features become predictive of significantly increased risk of claims. By plotting the Leaked Information score for each organisation against its relative impact on the organisation's expected claims frequency, the dependency between an increased score and increased risk can be visualised.

Figure 3: Relativities: Visualising the dependency between Leaked Information score and claim frequency risk

As with Orpheus' data overall, Leaked Information is more predictive for ransomware claims specifically than for claims in general. While higher scores can be associated with 1.25–2.25x higher likelihood of any claims, these indicate a 2–4x higher likelihood of ransomware claims.

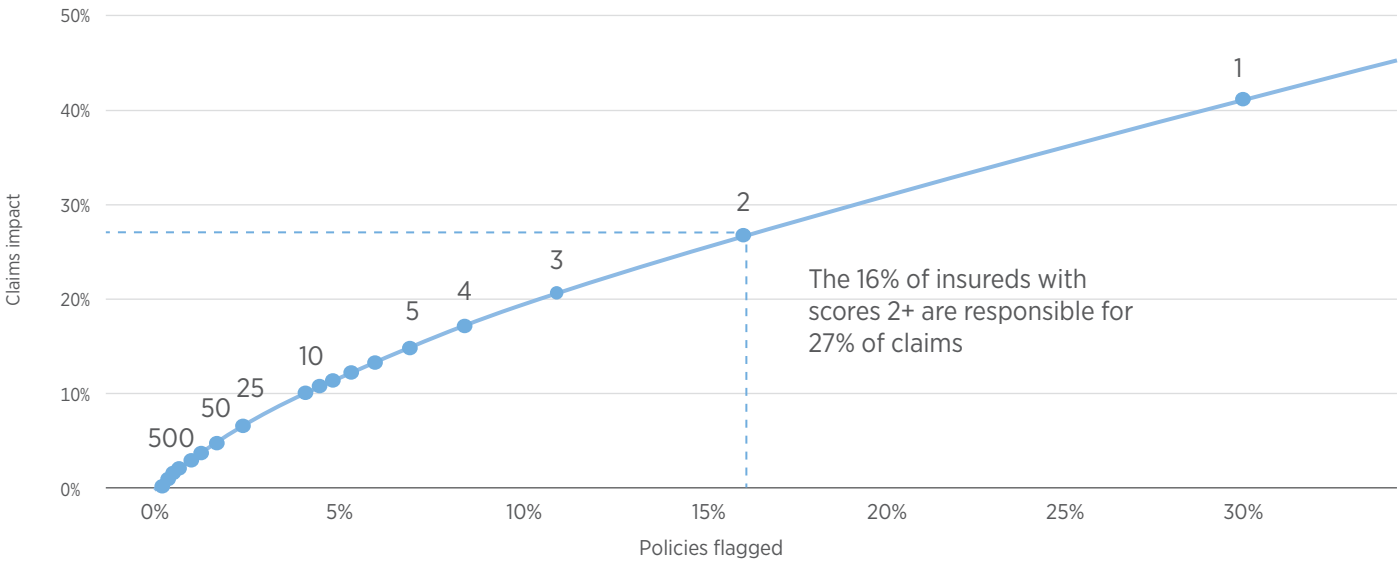
One of the key findings from our analysis is the consistency of leaked information's predictive power across companies of varying sizes. Whether a company falls into the nano/micro, small, or large revenue band, the Orpheus Leaked Information score remains a reliable indicator of potential claims. This consistency allows insurers to apply this metric across their overall underwriting approach.

Analysis of the distribution of Leaked Information Score and dependency data reveals interesting insights:

- (84%) organisations have a score of 0-1 on this metric, hence why these see relative claims likelihoods on or slightly below average.
- The minority of organisations with higher scores see elevated claims frequency risk: at scores of greater than 2, and more so at scores of greater than 6

To understand this relationship better, we assessed the relative trade-off between the proportion of insureds with particular scores, and the proportion of claims that these insureds represent. This allows us to evaluate the efficiency of setting different criteria for selection, pricing, and terms, particularly when insurers are competing in a soft market and are looking to minimise the friction at underwriting for insureds.

Figure 4: Trade-off between flagging policies and claims impact of Leaked Information score thresholds



For insurers who are looking to grow more aggressively, the potential impact of relaxing underwriting criteria — on both predicted claims frequency and applicant experience — can be assessed, while those looking to increase underwriting discipline can understand the extent to which this will flag insureds for review and the expected reduction in claims frequency from this change.

In both cases, insurers can then decide how best to respond to insureds that are posing increased predicted claims frequency — by avoiding selection, by adjusting pricing, or by setting terms that limit exposure to these risks. While this is not currently directly answered by claims frequency predictions, future studies incorporating severity will give better insight into pricing adequacy for expected claims.

Conclusion

We hope that this report adds to the broader understanding of how to assess and deploy external technographic data at underwriting.

Identifying relevant considerations, such as leaked information, which are proven to lead to claims can ensure that insurers are focusing efforts where these will have an impact and add value to insureds (who are also aiming to avoid incidents that lead to claims!). Additionally, understanding the distribution of these features and their impact on claims likelihood will hopefully lead to a more nuanced consideration of underwriting approaches — balancing competing objectives of growth and minimising risk.

The importance of leaked information for predicting claims frequency highlights for insurers the value of using this data source in assessing risk, especially when using threat intelligence to interpret and filter the raw data.

Future Gallagher Re studies will examine a number of the additional considerations raised from our research so far: the relevance of technical footprint as a predictor of claims, the importance of claims detail & tagging, and assessing claims severity to achieve a view of pricing adequacy.

Authors

William Gildea
Cyber Security Consultant
william_gildea@gallagherre.com

Mohit Motwani
Data Analyst
mohit_motwani@gallagherre.com

James Poynter
Global Data Science Lead
james_poynter@gallagherre.com

Learn more about our client-focused, collaborative approach.
Connect with us today at **GallagherRe.com**.

It's the *way* we do it.

Gallagher Re

GREGLOB104979