Kroll | A Division of DUFF & PHELPS

COGNISTX™

Leveraging Artificial Intelligence to Proactively Detect, Track and Minimize Data Breach Threats

EXECUTIVE SUMMARY

This article outlines at a high level how Kroll is advancing the art and science of artificial intelligence (AI) to help organizations proactively detect the occurrence and mitigate the severity of data breaches. Kroll is using machine learning to curate and extract meaningful information for organizations from the immense amount of data on the Deep Dark Web (DDW) by reducing and optimizing the search space, and enabling easy review to identify risk markers for data losses. Cognistx's data visualization tools help demonstrate the timeline of events on the DDW potentially leading to a data security breach based on exposure of PII or loss of other sensitive data. Ultimately, this insight can be used to help organizations proactively detect data losses from escalating into major events.

Deep Dark Web (DDW) Monitoring with AI

By Anju Chopra, SVP Cyber Technologies, Identity Theft and Breach Notification, Kroll; Heather Williams, VP Product, Identity Theft and Breach Notification, Kroll; and Dr. Eric Nyberg, Cofounder and Chief Data Scientist, Cognistx

Regulators, consumers and investors/stakeholders are increasingly not willing to accept the prevailing "not if, but when" defeatist attitude regarding data breaches. For example, the commission set up to oversee implementation for the European Union's General Data Protection Regulation (GDPR) unequivocally states, "As an organisation it is vital to implement appropriate technical and organisational measures to avoid possible data breaches."

So, we are not posing a rhetorical question when we ask: What if you could predict, with a great degree of confidence, where and when your data might be compromised?

For years, Kroll has been advancing the use of artificial intelligence and machine learning to not only discern data that has been lost, but also how to find detailed information in a noisy world. One of the key areas to find answers is in the deep dark web (DDW).

The DDW grows at a staggering pace every day across several protocols, forums and a multitude of sources. Kroll's proprietary technology, directed by highly experienced threat intelligence experts, has been identifying and indexing DDW data for more than 14 years. Our efforts have resulted in an extraordinary data resource, which is continually refreshed and curated. Most importantly, it enables us to help law enforcement, government and organizations in the public and private sectors understand not only existing exposure levels, but also the patterns and contexts that can potentially detect and mitigate data loss, be it trade secrets or personally identifiable information (PII).

Artificial intelligence and machine learning are critical for efficiently and accurately extracting meaningful information from the DDW's millions of files and petabytes[1] of information. In this article, we provide a high-level overview of how in partnership with Cognistx, Kroll has developed a comprehensive and intelligent solution that addresses the challenge for organizations in three phases:

1. **Reduce and optimize the massive search space of the DDW to better direct analytical focus**
2. **Find needles in this massive haystack, i.e., information that is pertinent to the organization**
3. **Discern patterns across the data that can serve as early warning system**

[1]One petabyte = 2^{50} bytes; 1024 terabytes, or a million gigabytes

REDUCE AND OPTIMIZE THE SEARCH SPACE WITH AI

At the start of every engagement, Kroll collaborates with the organization to create a list of key terms, IP addresses, domain names, etc., that are unique to the organization, what we call the Dynamic Signature Profile (DSP). Given the near-overwhelming vastness of the DDW, the first step requires that we optimize the search space, enabling us to separate vital signals from the noise.

First, we use AI to remove files that may contain DSP terms but which do not represent a risk. These can include items such as large pdf books, media interviews, marketing collateral, speaking engagements, etc. This is accomplished by training several supervised machine learning models to distinguish between pertinent and irrelevant files. Kroll and Cognistx have trained models that achieve 99.97% accuracy on this task by combining an initial dataset of several thousands of files with insights from highly experienced threat intelligence analysts. This allows us to reduce the search space significantly so subsequent AI-based analytics can focus on an optimized search field.

FINDING ORGANIZATION-SPECIFIC "HAYSTACK NEEDLES" IN THE DDW

After reducing the search space, the next step is to use cognitive clustering to find patterns across the reduced but still very massive data store.

Clustering approaches supplemented by interactive human judgment are fundamental to the Kroll-Cognistx solution. Over time, we have been able to organize files based on salient terms and metadata into coherent groups that reflect various aspects of exposure to potential data losses.

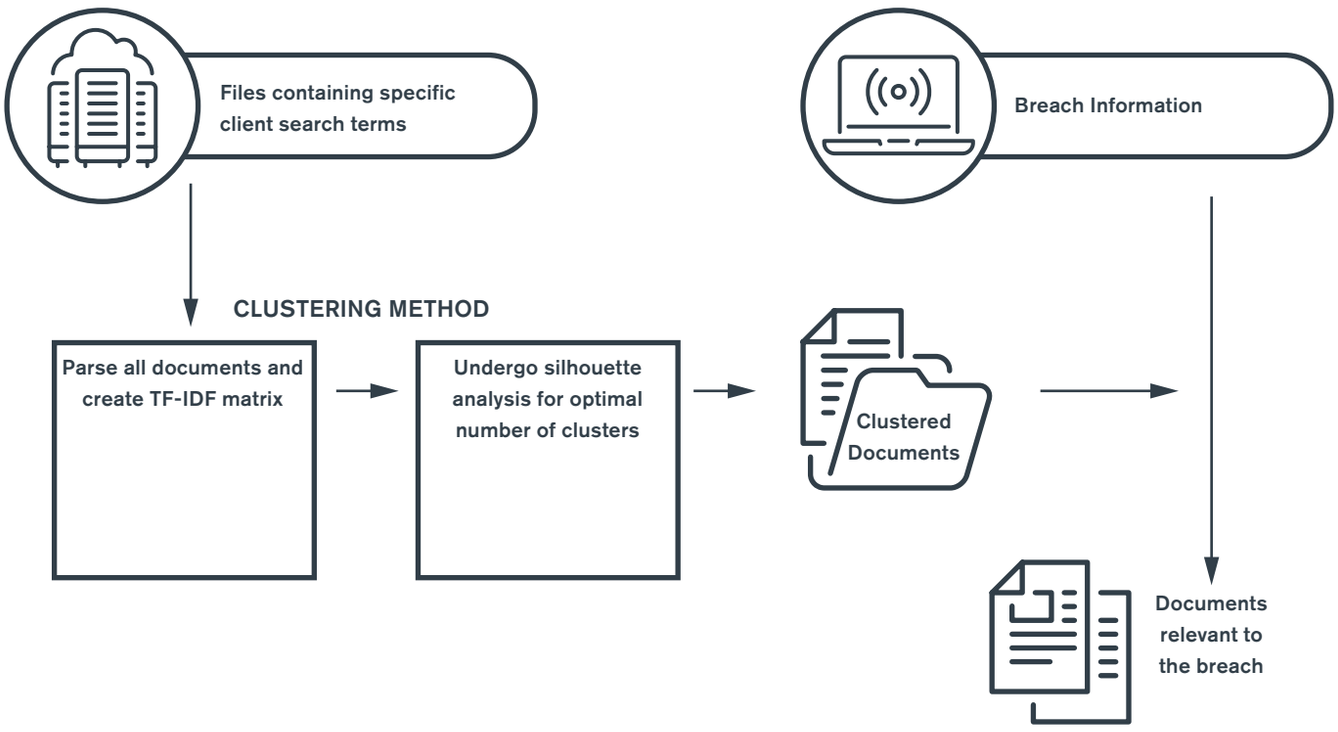
In this way, we have uncovered several productive patterns that can be routinely monitored by analysts using AI tools. This combination of man-machine analysis has helped us find several valuable exposure indicators across different industries. For example, in one case, we found large numbers of travel itineraries for an airline, which indicated a specific vulnerability for this organization.

Learning from the above analysis, we apply similar techniques to the relatively smaller set of files that are pertinent to a specific organization, enabling us to find risk indicators specific to them in Kroll's massive DDW datastore. As clusters emerge for a given organization, we apply our exposure assessment model to help determine the organization's level of risk on the DDW.

The methodology used for this phase includes developing organization-specific exposure indicators that take into account context idiosyncrasies. These indicators are built on the basis of signals mined from the text, which we use to create a "salient term matrix." We also examine the format and review occurrence characteristics, such as the timing, protocols and locations of organization-pertinent files.

An active learning loop with human analysts continues to refine and expand these signals. This loop also helps optimize the number of clusters for human review, which enhances our ability to find evidence pertinent to potential data loss events for an organization.

CLUSTERING & BREACH ANALYSIS: METHODOLOGY OVERVIEW



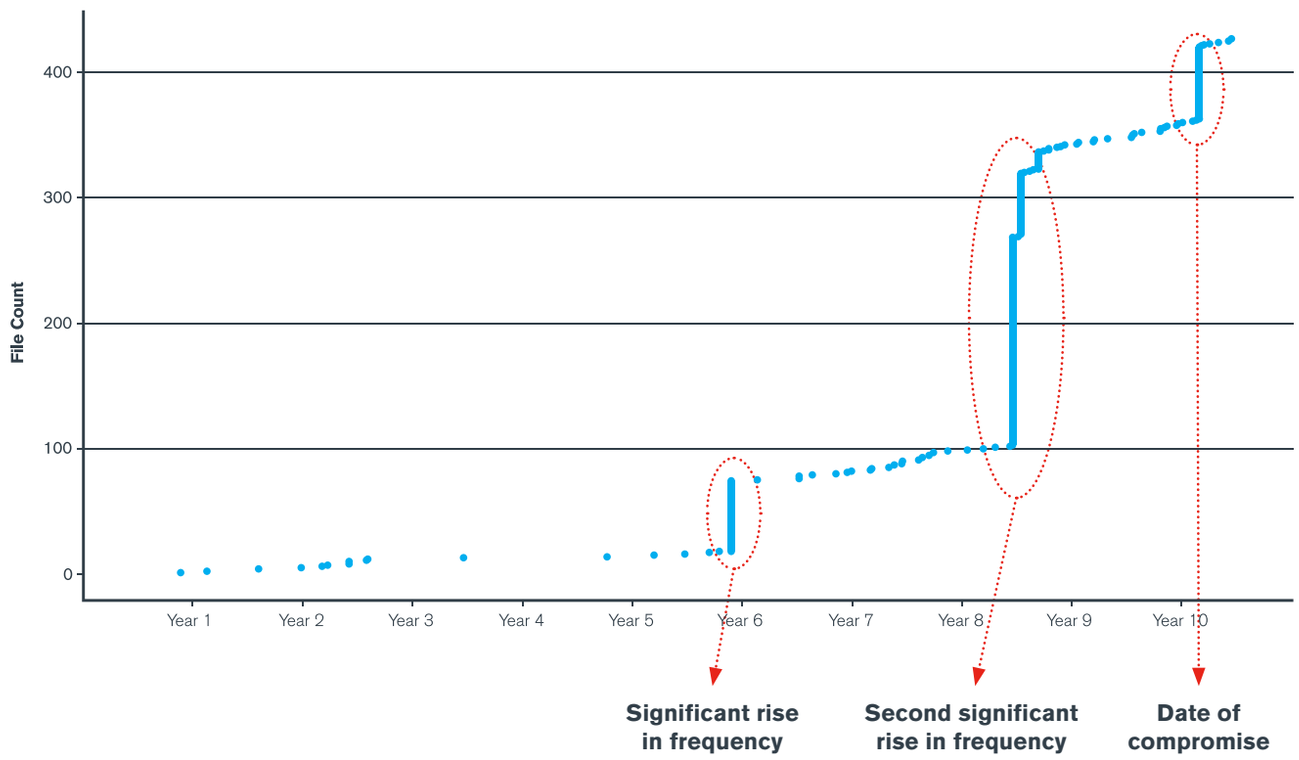
PROACTIVELY DETECTING DATA LOSS BASED ON FILES RELATED TO AN ORGANIZATION BEING SEEN ON THE DDW

Once we have determined organization-specific exposure indicators and clusters, the next phase involves a timeline analysis to study historical pre-breach patterns with the hope of eventually predicting potential breaches utilizing AI. The basic premise of the model is outlined below.

We have found that our clustering analysis shows a significant increase in the number of files on the DDW for an organization after a data loss event. We have also noticed that bad actors will accumulate files. Once a threshold of sufficient data is met, the information is exploited for potentially nefarious activities. For example, the chart below shows how clusters emerged for a given organization over a 10-year time period. The third data loss event in the past five years ultimately gave bad actors enough exposure to act on the data.

We have trained models that can detect the surge in activities for organization, which combined with active monitoring by human analysts, helps us to detect potential breaches. Our goal is that these automated systems and human team members can help by alerting organizations as these clusters emerge, enabling us to work together to proactively detect data loss before it becomes a larger incident and to help them understand where the data exposure is coming from within their organization.

FREQUENCY OF OCCURRENCE OF EMAIL/PASSWORD FILES LEADING UP TO THE BREACH



CONCLUSION

By integrating Kroll's investigative expertise with cutting-edge AI and machine learning from Cognistx, we are helping organizations better understand and address their data's exposure on the DDW. Better yet, our hope is that as we continue to refine and grow our models and methodologies, we will be able to help our clients proactively detect data loss and prevent those data losses from escalating into major events that can harm their operations, finances and reputations for years to come.

Kroll | A Division of DUFF & PHELPS

COGNISTX™

ABOUT THE AUTHORS

Anju Chopra, SVP, Cyber Technologies, Identity Theft & Breach Notification, Kroll

In a career spanning over 20 years, Anju has been a leader in delivering innovative, often ground-breaking advances in complex technology systems, cyber security, artificial intelligence and enterprise architecture. She has particular expertise in developing cyber security and identity theft remediation products that integrate artificial intelligence technology. Anju's strong business acumen and entrepreneurial vision have resulted in strategic solutions that have transformed client services and the internal operations that support them.

anju.chopra@kroll.com | +1 412.400.6120

Heather Williams, VP Product Management, Identity Theft & Breach Notification, Kroll

Heather has been with Kroll for over 12 years and a driving force for product innovation for nearly a decade. Her expertise in the fields of identity theft, breach response and cybersecurity have led to the development of enhanced solutions for these complex issues. She has been instrumental in developing dark web monitoring solutions as well as cyber investigative resources that integrate artificial intelligence technology to better serve clients and their customers.

heather.williams@kroll.com | +1.615.577.6715

Dr. Eric Nyberg, Co-founder and Chief Data Scientist, Cognistx

Dr. Eric Nyberg is a tenured Professor at Carnegie Mellon's School of Computer Science and has worked on a broad range of AI applications — automatic language understanding, translation, and generation, advanced information retrieval and ranking, and automatic question-answering systems — since the 1980s. As a member of the original Watson development team, he helped IBM develop a generalized, scalable architecture for multi-strategy question answering systems, as well as specific techniques. In 2015, he co-founded Cognistx, an applied AI company, building multi-strategy AI systems for clients across the U.S

ehn@cs.cmu.edu | +1 412.304.2978

About Kroll

Kroll is the leading global provider of risk solutions. For more than 45 years, Kroll has helped clients make confident risk management decisions about people, assets, operations and security through a wide range of investigations, cyber security, due diligence and compliance, physical and operational security and data and information management services. For more information, visit www.kroll.com.

About Duff & Phelps

Kroll is a division of Duff & Phelps, a global advisor with nearly 3,500 professionals in 28 countries around the world. Our clients include publicly traded and privately held companies, law firms, government entities and investment organizations such as private equity firms and hedge funds. We also advise the world's leading standard-setting bodies on valuation and governance best practices. For more information, visit www.duffandphelps.com.

© 2019 Duff & Phelps, LLC. All rights reserved. KR190032