Applying AI to secure the payments ecosystem

Cybercriminals, today, are well organized and resourced. Cybercrime is a $600 billion enterprise-strength industry —or almost one percent of global GDP that is growing at an unprecedented pace. As is evident from the below graphic (figure 1), data breaches continue unabated, and so does fraud, even though spending on cyber-security continues to grow. For merchants in the eCommerce space, where financials can be stolen via malware injected into the merchant websites, at scale, the threat posed by malicious actors continue to mount. Data breaches pose a unique challenge in the payment ecosystem as a single point of compromise can lead to loss of thousands and sometimes even millions of payment credentials. Thus, it is very important to identify and address these events as early as possible.

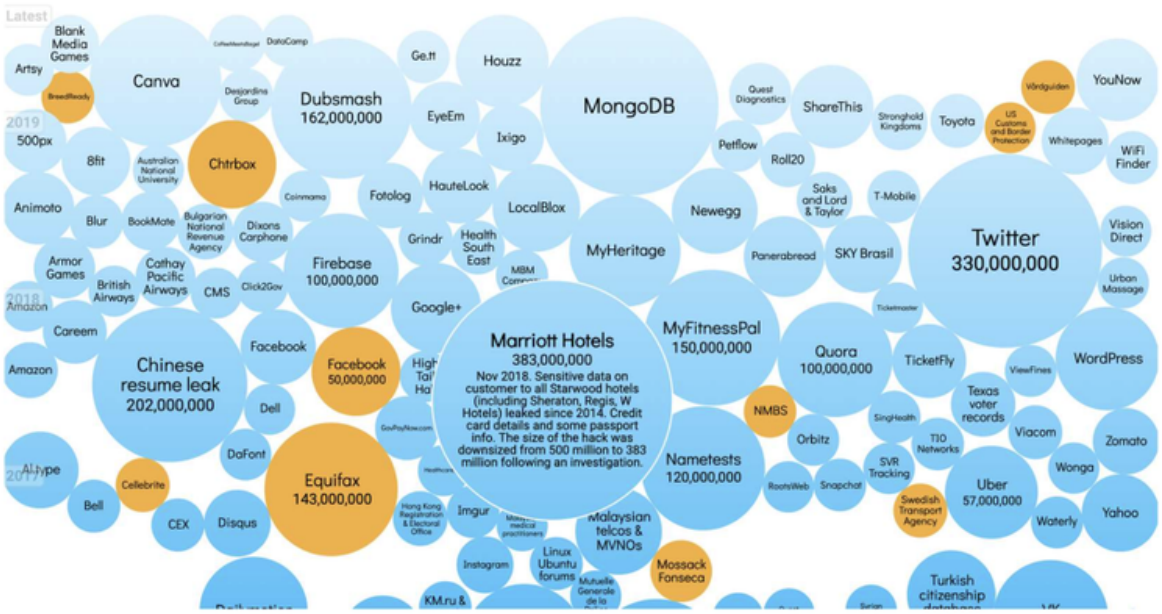

Figure 1. Infographics from informationisbeautiful.net showing known data breaches

Being a leader in digital payments, it is Visa's top priority to secure payments and protect our customers. As the network connecting XXB cards, **YYM merchants, and ZZ financial institutions**, Visa is uniquely positioned to detect data breaches affecting merchants in a timely manner. The problem of detecting data breach gets very challenging due to:

1. Limited information on past breaches making labeled data very rare
2. Varying nature, size, type and length of data breach
3. Early detection cannot wait for all signals to materialize
4. High accuracy requirement, as every investigation is a costly affair

To address these challenges, our team in Visa Research (VR) has looked into a variety of state-of-the-art approaches and identified a CNN based approach that alleviates these. Our solution not only detects data breaches that often go un-noticed, it also detects them faster thereby preventing more card info from being stolen as well as prevent monetization of already stolen payment credentials. However, as we know, CNN primarily works on image data and it's with images most of the benefits of a CNN approach can be reaped.

In order to apply CNN for breach detection, the first question that comes to mind is how do we convert payment data to a format suitable for CNN. To understand our approach, let's look at shopping journey of cardholders post visiting a merchant ABC as shown in figure 2. The horizontal axis here represents days while each point on vertical axis correspond to a different cardholder. We start with 4 cards A, B, C and D that visit this merchant on day 0, and their visit recorded via green dots. Post visiting this merchant, cardholders go on to shop at other merchants and their subsequent good visits are recorded as green dots while any risky/fraudulent visit is recorded via red dots. In this particular example, card A performs a bad transaction 60 days post shopping at ABC store.

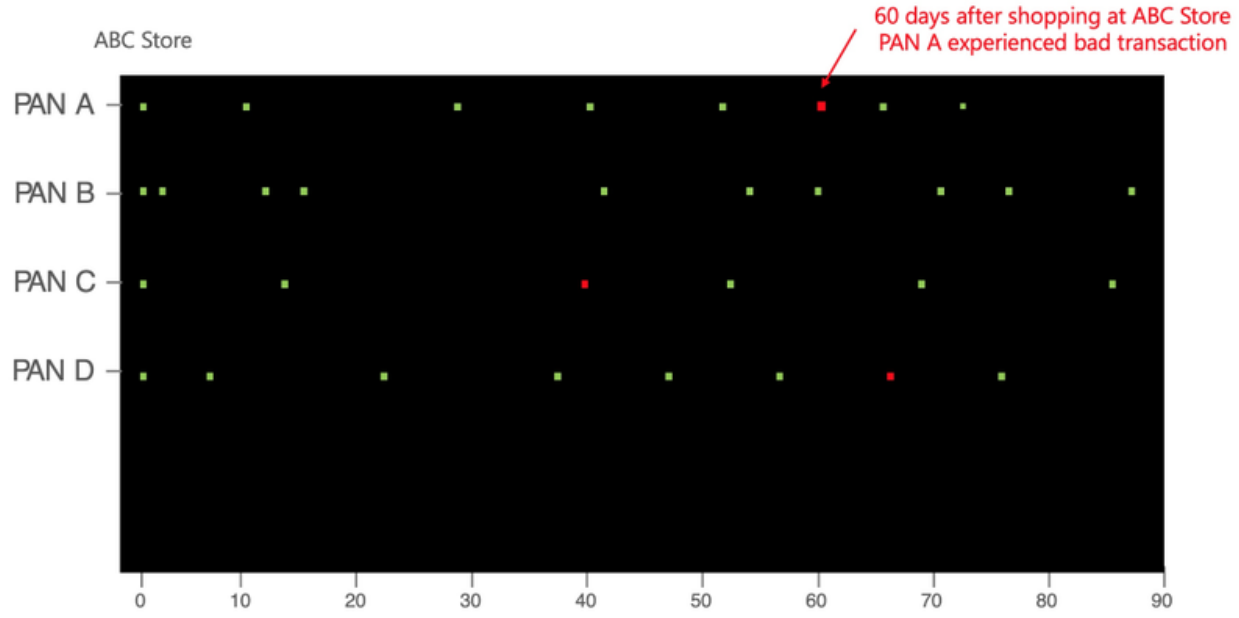

Figure 2. Shopping journey of 4 cards that visit ABC store

Generalizing this further with all the cardholders that visit this merchant we get an image as shown in figure 3. As expected, no specific patterns are observed for this as well as most of the merchants. However, when we look at this image for merchants who have been part of known data breach the story is quite different (figure 4). Here we see distinct patterns, indicating monetization or attempt to monetize stolen credentials. These indicative patterns can be effectively detected by a well trained CNN classifier. To summarize, the input data for our CNN model is image for each merchant derived by plotting the shopping journey of cardholders that visited this merchant.

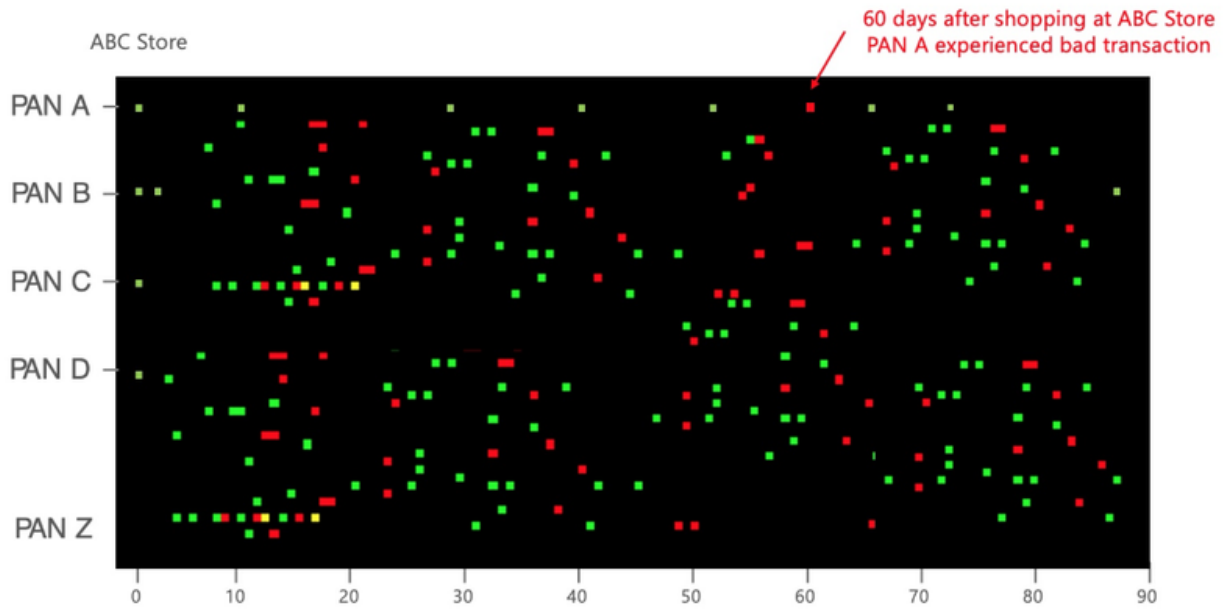

Figure 3. Image resulting from following shopping journey of all cards that visited ABC store

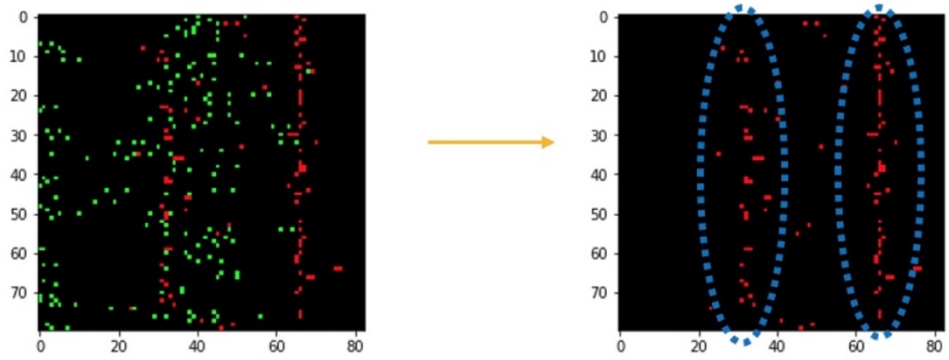

Figure 4. Image created for a known breach merchant showing distinct patterns indicative of data breach

As we recall, one of the biggest challenges with data breach detection is the scarcity of labels and varying nature of data breach. The problem gets further complicated by the fact that a big portion of data breaches never get reported leading to incorrect labels for those merchants. We have applied a two-stage algorithm based on CNN to address these challenges as shown in figure 5. During stage 1, label data is augmented. Label data is first augmented using image augmentation techniques (and hence justify our choice to use CNN for solving this problem) like geometric transformations, mixing images, kernel filters etc. and new labels are generated. Then a semi-supervised CNN is used to further augment the label information and create more labels

on the vast unlabeled merchant data. In stage 2, augmented label information is used along with static features to train a deep learning model for early breach detection.

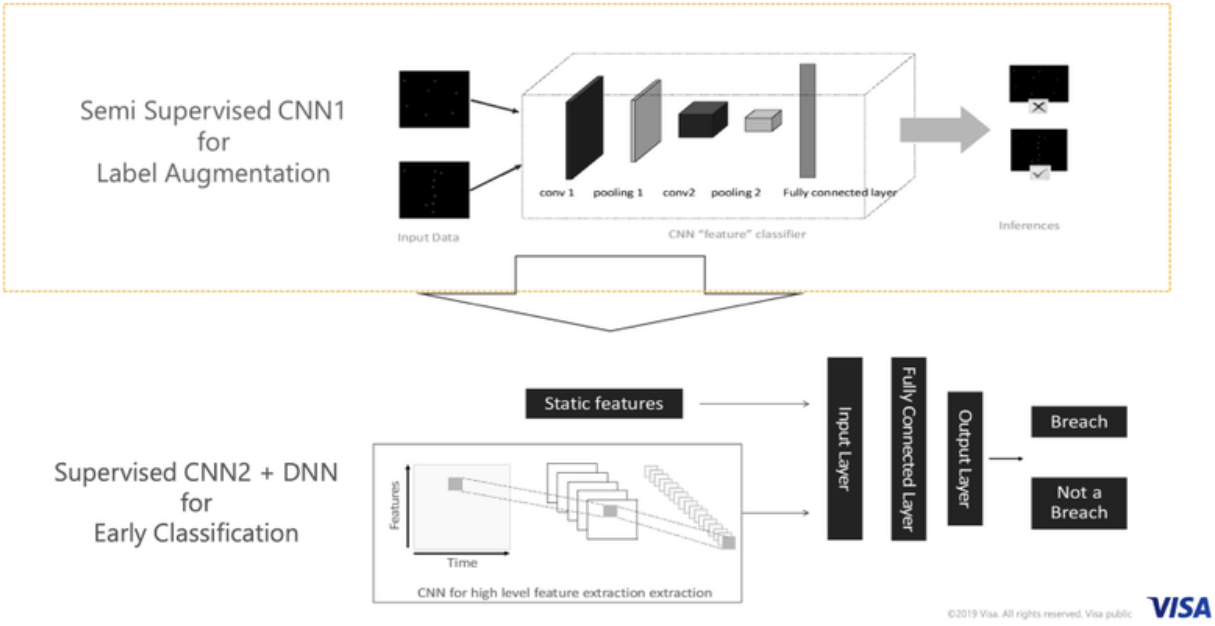

Figure 5. Two step breach detection model architecture

Using this approach, we were able to get more than a 200% lift in detection capability compared to traditional ML approaches using just hand-crafted features. Not only was our approach able to detect more breaches with a higher 1:1 accuracy, it also detected breaches on average 3 weeks prior to them being detected using traditional methods. As part of the Applied Prototyping team in Visa Research, in addition to building the model, we have created a fully functioning prototype called Breach Identification Tool (BIT) that is being used by Visa Investigators to research and surface new breaches and protect the payments ecosystem.

In terms of future research, we are investigating using GANs to supplement or replace the semi-supervised stage 1 process for label augmentation. Our team is also investigating the prospect of building an intelligent system that identifies the likelihood of fraud on a card involved in a breach. This will help issuer to proactively reissue cards with high likelihood of fraud, while more tightly monitor others with moderate to low chances of fraud. Overall, our research in this domain is helping Visa stay on top of fraudsters and prevent the payment ecosystem from their attacks.