Money-rity Report: Using intelligence to predict the next payment card victims

Clare Gollnick
Chief Technology Officer
Terbium Labs

Cathal Smyth
Machine Learning Researcher
Royal Bank of Canada

Danny Rogers
Chief Executive Officer
Terbium Labs

Kory Fong
Head of Cyber Security Research
Royal Bank of Canada

***Abstract*—Carding is a big business for criminals and a big liability for the payment industry. Losses from credit card fraud amount to more than $10 billion dollars annually. The money paid to hackers and gangs that supply large online carding markets feed directly into a criminal economy, enabling not only further financial fraud, but also terrorist activity. Despite the high stakes, combatting credit card fraud historically has been a slow, reactionary process. There is a better way. By combining intelligence gathered from online carding forums with transactional data, we demonstrate an approach that can predict future victims before fraudulent credit card transactions occur.**

Keywords— dark web, intelligence, carding, common point of purchase, fraud prevention, predictive analytics

I. Introduction

Technological advances do not discriminate by integrity or intent. The internet has enabled the rapid growth of businesses online, breaking down state and country borders, aiding in the globalization of business. The same is true for criminal economies. A booming financial fraud business, centered around stolen credit card data, flourishes on the dark web. These markets exist almost outside the reach of the law enforcement. Criminal liability is complicated, often requiring coordinated efforts among the governments of multiple countries. Taking a cue from intelligence agencies, thought leaders in the payment industry are looking for new and novel approaches to combatting fraud: data intelligence.

Fraud prevention analytics almost always begins with labeled transaction data. A transaction or attempted transaction may be categorized by a heuristic or machine learning algorithm at the point of sale, ideally before merchandise is transferred or services are rendered. Institutions also rely on client self-reporting or reporting from third-party services to begin an investigation. Once a transaction is confirmed to be fraudulent, it is added to a pool of other recent fraudulent transactions. Transaction history from these cards are then cross-referenced in a process known as common-point-of-purchase (CPP) analysis. These CPPs form a shortlist for a possible point-of-compromise (POC) locations. Mitigation efforts begin here, including card re-issuance, risk monitoring, and notification of law enforcement.

Labeled transaction data is valuable, but comes with substantial cost. In order for an investigator or researcher to start an investigation, a fraudulent transaction (or attempted transaction) must occur. With this delay, monetary loss is ceded as a necessary initial step. More importantly, by the point a stolen card is used, criminals have already turned a profit selling hacked card information, enabling the ecosystem of stolen cards to move on to the next breach as soon as mitigation efforts begin. The best solution to this problem would prevent fraudulent transactions, but also inhibit the profitability of large-scale online carding forums. Here, we present a method that inverts the uncertainty of the traditional POC analysis. Rather than starting with certainty of the victims (labeled fraudulent activity) and having uncertainty as the number of sources, we use a collection of uncertain potential victims assumed to be arising from a single POC.

II. Methods

A. Building Candidate and Control Groups

The intelligence-centered approach to combatting fraud starts with the private and anonymous collection advertisements on online carding forums and paste sites. The advertisements do not reference the source or method by which the cards are acquired. This information is tradecraft and is kept secret. Instead, each advertisement provides information the malicious consumer would need to execute fraudulent transactions: a Bank Identification Number (BIN), a price, etc. This extra metadata acts as a (sparse) but informative profile of the victim. We start with identifying a group of potential victims for each listing, from within an issuers known customer list. The set of potential victim accounts for each listing is referred to collectively as a 'candidate group'. A candidate group may be anywhere from 1 to 1000 or more accounts. Fig. 1 depicts this process that eventually leads to the successful detection of a POC. In this example, a hypothetical gym named "XSS-Fit" is the sole point-of-compromise for the advertisements on the market. For each potential victim account, a transaction history and customer metadata are collected. The complete set of potential victims within all candidate groups helps to define a joint distribution of characteristics (age, location etc.) that profiles the potential victims of fraud. Using this joint distribution, a random sample

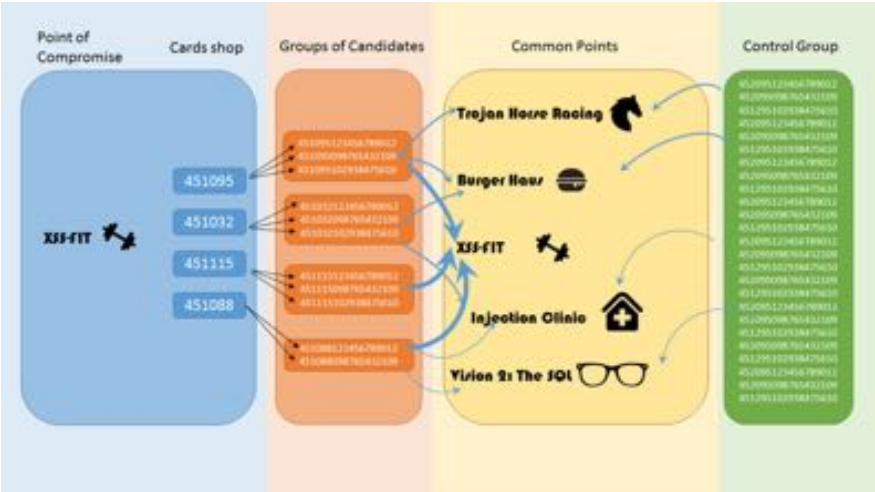

Fig 1 An illustration of how dark web intelligence from carding markets can be used to identify common points of compromise.

of card holders that have similar characteristics, but outside any of the candidate groups, is selected to serve as a control.

B. Point of Comprimise (POC) Analysis

A Bayesian common point-of-compromise detection approach based on the BreachRadar algorithm developed by Araujo et al. [1] was used for this study. To be successful, an identified POC must be localized in time as well as in space (physical or digital). A combination of spatial information such as merchant name, merchant ID and terminal ID, as well as temporal information such as year and week, can serve as initial hypotheses for a POC. For simplicity, we will refer to this combination as a "merchant". However, this approach generalizes; for example, if the suspected POC was a conglomerate of different clothing outlets, one could first focus on the Merchant Category Code (MCC) or use of a specific point-of-sale system.

Briefly, the BreachRadar algorithm employs a bipartite graph network model. Credit accounts and merchants form two non-overlapping sets of nodes. Transactions are modeled as graph edges (a link from accounts to merchants). Edge weights carry blame from fraudulent cards to potential compromised merchants. In the original paper, fraudulent transactions labels were certain, therefore weighted strongly and equally. In the current approach, however, we are uncertain as to whether our labels are accurate (potential victims instead of known victims) so we assign less blame and with less confidence accordingly. For example, one can encode confidence in labels through the weighting of the transactions in the edge matrix N_{ij}. Alternatively, this information can be captured by the blame parameter as shown in equation (1). If the client is a part of a group of 1000 candidates, their blame is 100 times less reliable than a client that is part of a group of 10.

$$B_{ij} = \frac{1}{N_{m(c-i)}} \frac{N_{g(m=j)}}{N_c} \quad (1)$$

The elements of the blame matrix are populated as above, where i and j represent the client and merchant index, respectively. N_m represents the number of merchants visited by potential victim i, N_g is the number of groups that shopped at merchant j and N_c is the number of candidates in the group that contains potential victim i. We assume a prior in the form a Beta distribution *Beta(0.2, 15)* that encodes a strong assumption that CPPs are not likely to be compromised.

After blame is cast, the posterior probability distribution (θ) can be updated as described in (2).

$$E[\theta_j] = \frac{\sum B_j + \alpha}{n_j + \alpha + \beta} \quad (2)$$

The blame matrix can then be updated from the posterior probability (θ) distribution as in (3).

$$B_{ij} = \frac{\theta_j}{\sum_{k \in N_i} \theta_k} \quad (3)$$

This repeats until the posterior probability distribution stabilizes.

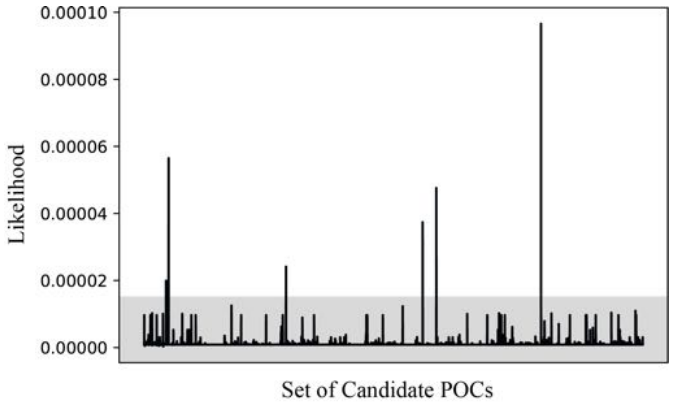

Figure 2 Relative Likelihood of different POCs after fifteen iterations of the BreachRadar algorithm. Only a few POCs rise above noise (qualitatively depicted by shading)

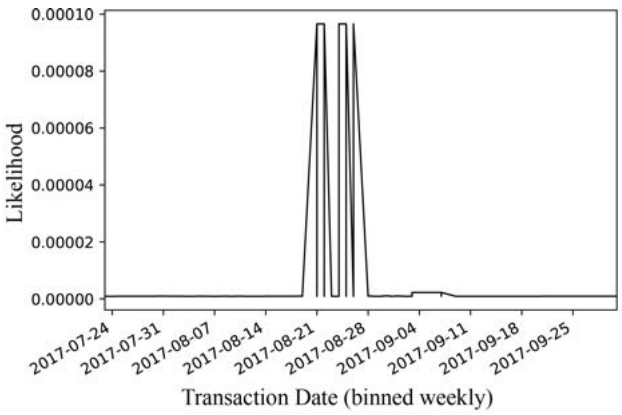

Fig. 3 Localizing the POC in time. Maximum likelihood POC, "Bob's" plotted overtime shows the timeline of the breach.

III. CASE STUDY

The following case study discusses a POC that occurred at a restaurant chain in Canada. The victims' cards were in fact spread over nearly a dozen bases and impacted a significant number of clients. In figure 2 we see the distribution of POC probability Θ after 15 iterations of the Breach Radar algorithm. A very strong signal is present for the merchant in question, which is henceforth known as "Bob's". The POC probability of Bob's is 15 times larger than random chance, and twice as large as the next value. As this model incorporates an increased level of unreliable and uncertain evidence, posterior probabilities are lower than with more reliable labeled data. However, relative patterns in the data are generally conserved.

Given an increased level of noise and uncertainty, supporting analyses are helpful in assessing the consistency of suspected POC with all available evidence. For example, Fig 3 localized this POC breach in time, by plotting the weekly probability Θ specific to the merchant Bob's. As we can see, the POC probability is essentially zero until the breach event in late August where it repeatedly spikes to .0001%. These repeated spikes occur in quick succession. The temporal correlation of candidate purchases across groups is a strong indicator that this identified POC is the actual source of the card advertisements.

In our implementation of this algorithm, a single advertised card is represented by a group of candidate cards. We can test the feasibility of this assumption on POCs identified by the CPP analysis. Figure 4 plots a timeline of purchases made at Bob's. The curve "Candidate Transactions" tracks the number of individual transactions or purchases made that week from cards within candidate groups, while "Group Spread" charts the number of candidate groups that have interacted with the merchant in a given week. A strong indicator that a POC is a true source of card advertisement is overlap of these two curves (see in Figure 4a). In contrast, we show in Figure 4b, a plot of the same values for the second most-likely merchant. In this case the merchant is a globally popular digital store. Note that Group Spread is broader across the timeline, but is consistently lower in value that "Candidate Transactions", indicating that many more candidates than groups shopped at this merchant, making it less likely as a possible POC.

IV. CONCLUSIOINS

The advent of the dark web has led to convenient, one-stop shops for the sale of sensitive information such as credit card details. Payment card fraud is the slush fund that underlies global criminal threats, from organized crime to global terrorism, in large part because of antiquated, reactive techniques and a dearth of innovation to more proactively combat it. Previously, discovery of the point of compromise has been limited to either to traditional fraud detection methods, or more controversially, through the purchase of client details. Neither of these approaches are ideal, as at best it leads to the continued flow of money into these black markets. In contrast, our approach represents a paradigm shift in fighting payment card fraud; by using dark web market intelligence combined with transaction data to predict both fraudulent charges and points of compromise, we can intervene before any loss occurs, stopping payment card fraud dead in its tracks and eliminating a major source of funding for the global criminal ecosystem.

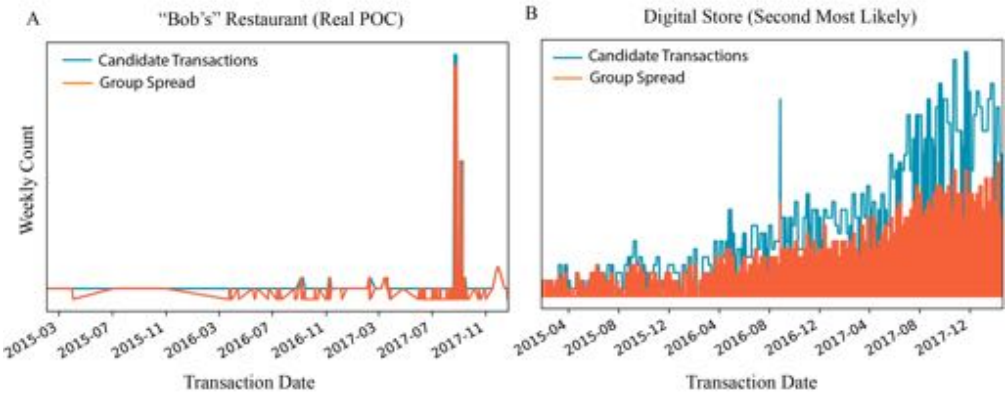

Fig 4 Additional evidence for the identity of the POCs include considering the temporal coincidence of the candidate transaction and candidate group level activity. A) High temporal overlap between group spread and candidate transactions for the true POC B) A false-positive POC shows no clear overlap between group spread and candidate transactions.

This form of POC detection should be relatively simple to implement for financial institutions with access to dark web intelligence. The detection could also be significantly augmented by collaboration across financial institutions, perhaps by using a form of differential privacy [2]. Global breaches can be discovered and jointly corroborated almost instantly after any attempt at mass monetization.

ACKNOWLEDGMENT

C.S. would like to thank Daniel Swerdfeger for his invaluable subject matter expertise within the domain of fraud investigations.

REFERENCES

[1] M. Araujo, M. Almeida, J. Ferreira, L. Silva, and P. Bizarro, "BreachRadar: Automatic Detection of Points-of-Compromise," Proceedings of the 2017 SIAM International Conference on Data Mining, 561-569

[2] Friedman and A. Schuster, "Data mining with differential privacy," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10, 2010, p. 493.