splunk> .conf2017

Payment Cards and Risk

How to detect stolen cards, pinpoint suspicious merchants and uncover compromised payment terminals

Gleb Esman | Sr. Project Manager, Anti-Fraud, Splunk
Felipe J. Hernandez | CEO, VPNet, Inc.

September, 2017 | Washington, DC

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Gleb Esman, Bio

1990's: **Anti-virus research and development**:
Belarus, Israeli anti-virus research and development.

2000's: **IBM T. J. Watson Research Center**, NY. Anti-virus development.
Heuristic virtual machines to detect known and unknown computer viruses and malware.

2000's-2010's:
Architecting and software engineering work in space of **e-commerce, cryptocurrency, payment processing and digital information management** solutions.

Before Splunk, till July, 2015: **Morgan Stanley**.
Working on data analytics solutions for financial services as well as helping to build Splunk-based security and anti-fraud applications.
Leading an effort to leverage Splunk as an anti-fraud platform for online banking.

Since August, 2015 – Sr. Product Manager at **Splunk**,
Anti-Fraud Products, San Francisco.

Author of several **Patent Applications for fraud detection with Deep Learning**.

splunk> .conf2017

Splunk Platform for Anti-Fraud

Why Splunk is the right fit to address challenges with sophisticated fraud?

- Splunk platform acts as the data driven central nervous system of organization.
- Splunk aggregates raw data coming in from multiple disparate sources and is indexed in real time.
- Data contains traces of anomalous behavior and patterns of suspicious activity.
- Advanced analytics and machine learning are utilized to effectively reduce exposure to fraud or loss

splunk> .conf2017

Case: Predicting and Preventing Chargebacks

Leveraging Splunk Machine Learning Toolkit to Predict Chargebacks on Credit Card Transactions

splunk> .conf2017

Chargebacks == EVERYONE is UNHAPPY

Intend to protect consumers from unauthorized transactions

- **Long** - Funds withheld from business until everything clears
- **Messy** - Chargeback resolution involves lots of paperwork
- **Expensive** - % processing fee + $10-25+ per case for merchant *regardless*
- **Long** - Takes 60-90 days to resolve
- **Messy** - May involve further arbitration between merchant and banks

'I Didn't Buy That': Friendly Fraud Costs Retailers $11.8 Billion a Year

Ecommerce will lose $6.7 billion in 2016 to fraud, according to data from eMarketer and LexisNexis.

Chargebacks Problem

Merchants are on the hook to lose

- Payment facilitators recovers chargebacks from **merchants**
- Issuing bank recovers the funds from the **merchant's bank**.
- Merchant bank recovers the funds from the **merchant**.

PROBLEM:

Merchants are kept aside and are notified when it's too late in the process to prevent penalties, fees and losses of funds and goods.

What Do We Need To Do To Predict Chargebacks?

Large Online Retailer approached Splunk for help

- Need to be able to detect transactions with high probability of chargebacks and put these through extra scrutiny.
- Need to be able to detect that in time close to actual transactions.
- Available data set to learn from was limited to log of 100,000 transactions
- Had confirmed record of only ~100+ confirmed chargebacks
- CC Transactions data contains about ~50 fields describing each transaction

Credit Cards Transaction Data

Some of the available data fields within e-Commerce Transaction logs

- Date/time
- Transaction Value: $ Amount
- IP Address (+city, zip)
- Customer Email(s)
- Shipping Address(es)
- Billing Address (es)
- **CaseStatus**: **Chargeback** / **Other**
- …
- Customer name
- **Risk Score**
- Customer ID
- Session ID
- Case Status
- Case Extra Data
- Phone number(s)
- …

Chargebacks: Machine Learning Solution

Splunk Machine Learning Toolkit delivers new SPL commands to apply variety of Machine Learning concepts to your data

https://splunkbase.splunk.com/app/2890/

Machine Learning is an ability of computers to learn and do predictions from data without being explicitly programmed

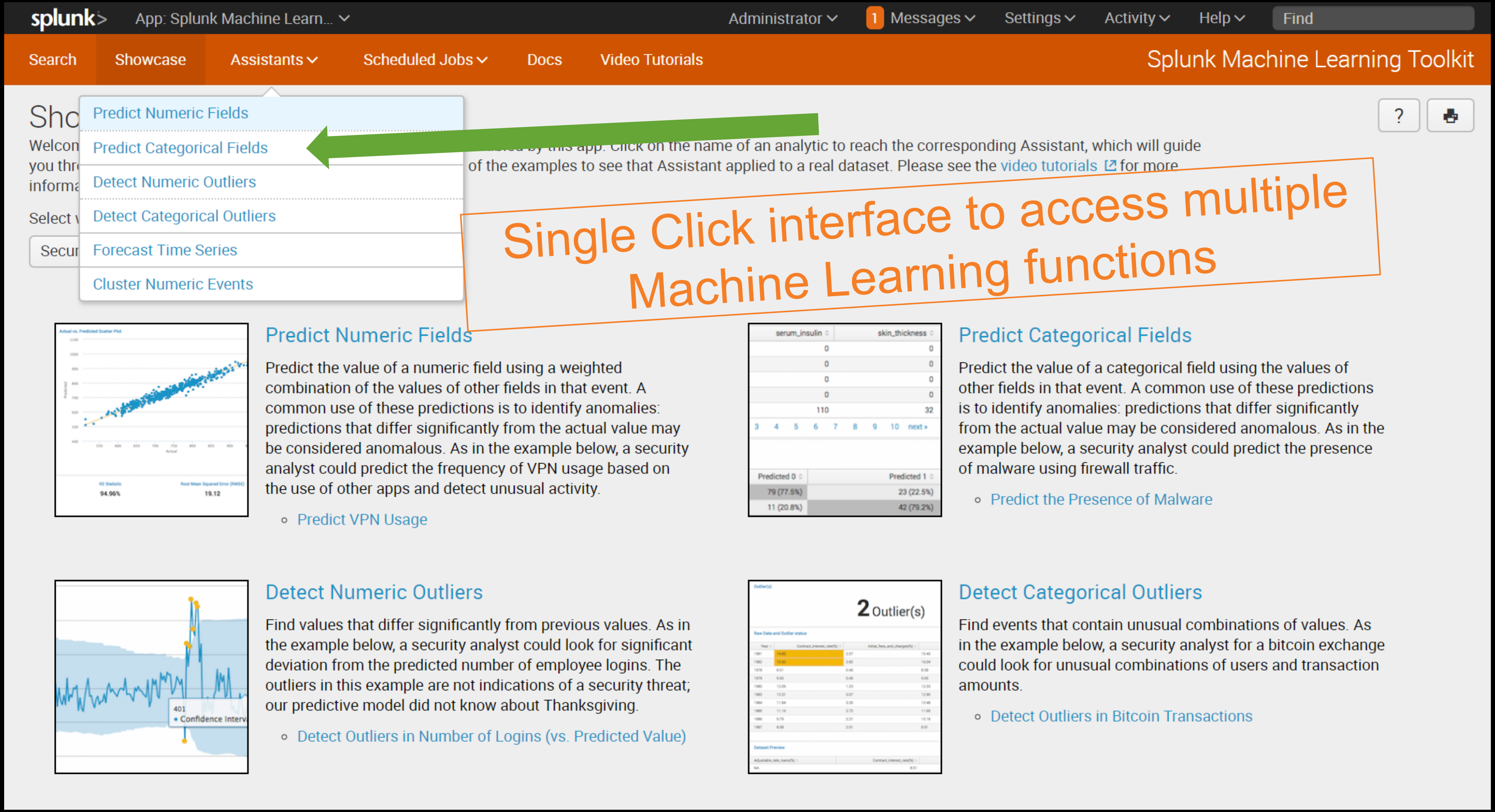
Building Chargebacks Prediction Model

Leveraging Splunk Machine Learning Toolkit

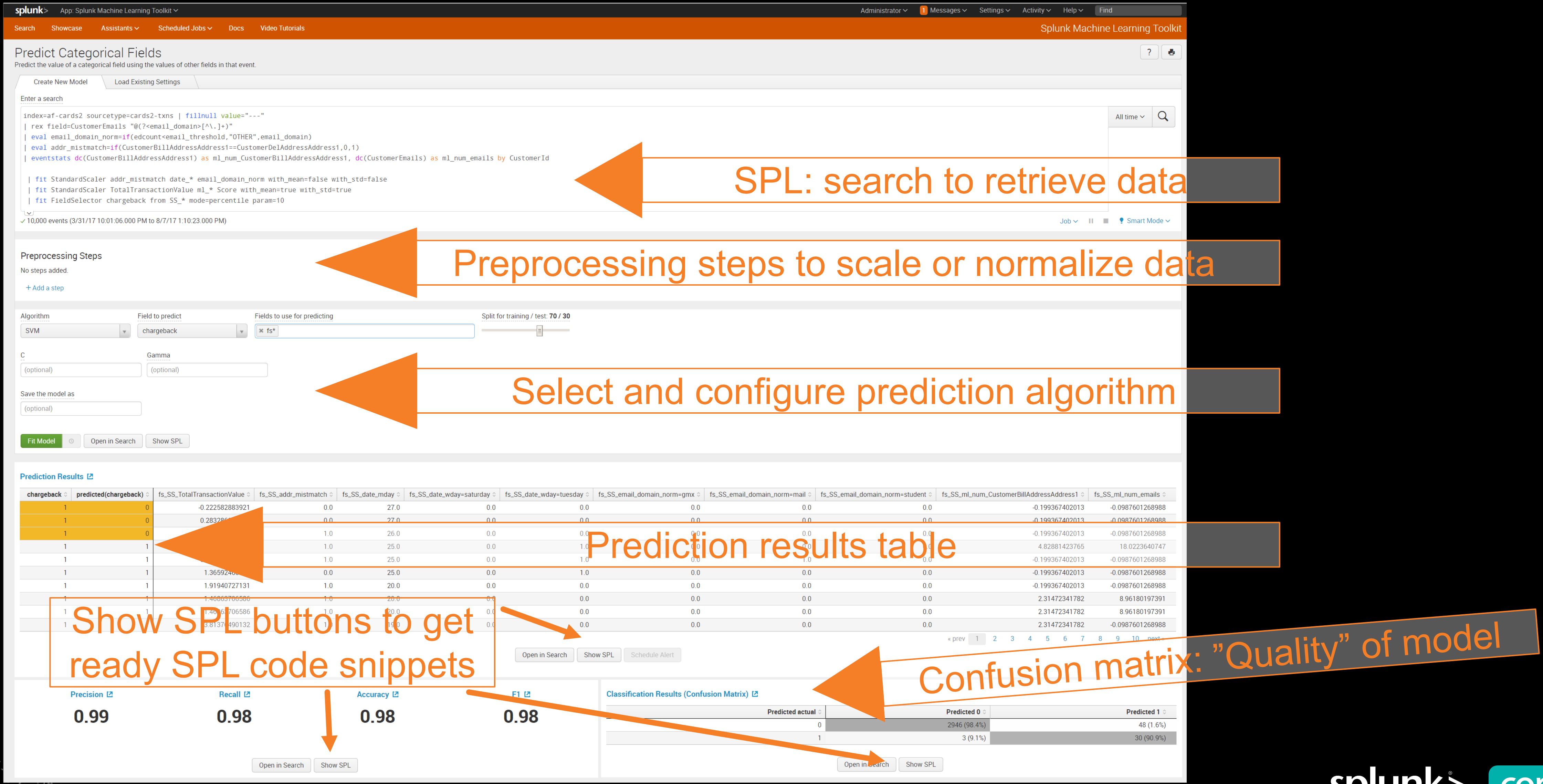
MLTK Benefits:

- Simple to use. Become data scientist in an hour!
- Web based interface to apply machine learning to your data.
- Guided navigation
- Guided assistants to build models on top of your data without coding skills
  - Predict Numeric and Categorical fields
  - Detect Numeric and Categorical outliers
  - Apply supervised and unsupervised learning techniques to solve problems
  - Detect unknown unknowns to catch attackers and fraudsters

Splunk Machine Learning Toolkit



Working with Splunk Machine Learning Toolkit



splunk> .conf2017

Building Chargebacks Prediction Model, cont.

General Steps

1. Devise SPL search to retrieve data
2. Select features available within data that may be useful
3. Consider preprocessing steps – everything needs to be numbers or converted to numbers. Many algorithms provided out of the box.

Secret Sauce:

1. Right Features (collected, extracted, engineered)
2. Right Data Preparation (scaled, normalized)
3. Right algorithm to train the model.
4. Right algorithm parameters, train/test ratios, data volume.

Secret Recipe To Devise A Good Model

1. Extract all possible features that may help to predict chargeback:
   - Static features (txn amount, email domain, address mistmatch)
   - Historical, behavioral and aggregate features (avg. txn, min, max, sequences, patterns)

2. Normalize categorical or “wildly” numerical fields:
   - … | StandardScaler email_domain with_mean=false with_std=false
   - … | StandardScaler txn_value other_* with_mean=true with_std=true

3. Apply Splunk MLTK “Magic” to pick only the best features:
   - Too many features hurts model predictive ability and slows down work.
   - Too many features cause model overfitting
     (ability of model to make correct predictions on unseen data)
   - … | analyzefields classfield=chargeback
   - … | FieldSelector chargeback from SS_* mode=percentile param=10

Actual SPL Used To Extract Features Of Transactions

```
index=af-cards2 sourcetype=cards2-txns
| fillnull value="---"
| eval chargeback=if(CaseStatus="ChargebackFraud" OR CaseStatus="ChargebackOther",1,0)

| rex field=CustomerEmails "@(?<email_domain>[^\.]+)"
| eval email_domain_norm=if(edcount<email_threshold,"OTHER",email_domain)

| eval addr_mistmatch=if(CustomerBillAddressAddress1==CustomerDelAddressAddress1,0,1)
| eventstats
  dc(CustomerBillAddressAddress1) as ml_num_CustomerBillAddressAddress1
  dc(CustomerBillAddressAddress2) as ml_num_CustomerBillAddressAddress2
  dc(CustomerDelAddressAddress1) as ml_num_CustomerDelAddressAddress1
  dc(CustomerDelAddressAddress2) as ml_num_CustomerDelAddressAddress2
  dc(CaseId) as ml_num_CaseId
  dc(DecisionResult) as ml_num_DecisionResult
  dc(SessionId) as ml_num_SessionId
  dc(ip) as ml_num_ip
  dc(CustomerEmails) as ml_num_emails
  dc(IPCity) as ml_num_ip_cities
  dc(IPPostalCode) as ml_num_zips
  by CustomerId
 | eval ml_len_phone=len(CustomerPhoneNumber)
 | eval ml_len_de2=len(CustomerDelAddressAddress2)
 | eval ml_len_ba2=len(CustomerBillAddressAddress2)

 | fields - date_second date_minute date_month
 | fields chargeback addr_* ml_* email_domain_norm IPCity IPPostalCode TotalTransactionValue date_* Score

 | fit StandardScaler addr_mistmatch date_* email_domain_norm with_mean=false with_std=false
 | fit StandardScaler TotalTransactionValue ml_* Score with_mean=true with_std=true

 | fit FieldSelector chargeback from SS_* mode=percentile param=10
 | fit SVM chargeback from fs_*
```

- Load data
- Set field chargeback
- Extract user email address domain
- Extract address_mistmatch feature
- Extract dozen of other features
- Exclude unrelated fields
- Standardize / normalize inputs
- Extract most important features
- Fit model

MLTK SPL Code To Predict Chargebacks

```
index=af-cards2 sourcetype=cards2-txns

  . . . . . .

  | fit StandardScaler addr_mistmatch email_domain_norm with_mean=false with_std=false
  | fit StandardScaler TotalTransactionValue ml_* Score with_mean=true  with_std=true

  | fit FieldSelector chargeback from SS_* mode=percentile param=10
  | fit SVM chargeback from fs_*
```

- StandardScaler – normalize data for prediction algorithm
- FieldSelector – automatically select only 10% (param=10) of the most important features carrying maximum predictive qualities for the target category
- SVM – chosen algorithm to predict chargebacks

Secret Sauce to Predict Chargebacks

Splunk + Machine Learning Toolkit results achieved with SVM model:

Accuracy of predicting good transactions: 98.4%

(Confusion Matrix)

Predicted actual	Predicted 0	Predicted 1
0	2946 (98.4%)	48 (1.6%)
1	3 (9.1%)	30 (90.9%)

Accuracy of predicting chargebacks: 90.9%

Conclusion

What helps to build successful model to predict chargebacks?

- Extracting relevant features for the prediction task is important. Ex: email is not important, however email domain is.
- Properly normalizing features (via StandardScaler and other algorithms) is important
- Automatically selecting only the best features (6-10% out of all available). Throwing away least performing features helps to minimize overfitting.
- FieldSelector is one of the great commands to automate field selection.
- RiskScore – third party input from risk calculation service did not carry any predictive value to improve chargeback detection.

Conclusion, cont.

Which features are important in predicting chargebacks?

- Out of 70+ selected features – these are the ones that were automatically extracted by FieldSelector as most influential to achieve the best results in classifying transaction as chargeback:
  - Address mismatch – billing and shipping addresses are different.
  - Total Transaction Value.
  - Email domain (Gmail, AOL, Yahoo, Hotmail, etc…)
  - Number of different billing addresses used by the customer
  - Number of different email addresses used by the customer

Case: Detecting Stolen Cards, Suspicious Merchants And Compromised Payment Terminals

Leveraging Splunk Enterprise and Splunk Machine Learning Toolkit to detect suspicious activity and fraud

Case study: Splunk Customer: VPNet

VPNet
SECURING EVERY NETWORK

- Leaders in Cybersecurity for retail customers in Puerto Rico.
- Serving 65% of the credit unions in Puerto Rico.
- Leaders in Cybersecurity for the Healthcare market.
- VPNet offers innovative solutions in the telecom and IT security industry.

Detecting Fraudulent Activity In Payment Cards

Leveraging Splunk Enterprise to detect suspicious transactions

VPNet approached Splunk for help in detecting fraud events within debit and credit card transactions. Limited transaction logs data was provided:

- Transaction Date/Time
- Card number
- Merchant location data (State)
- Transaction type (Purchase, Cashback)
- Merchant name
- Transaction Amount
- Terminal ID

\+ List of known compromised card numbers

Payment Cards transaction log:

```
2017-03-30,03:35:57,CARD01060092[redacted],PR,PURCHASE,$41.36,456560,1441456560,HAT[redacted]
2017-03-30,05:29:00,CARD01053505[redacted],PURCHASE,$81.00,059820,1664059820,HATHKI[redacted]
2017-03-30,05:34:59,CARD01060060[redacted],PR,PURCHASE,$169.26,441228,1266441228,HA[redacted]
2017-03-30,05:40:28,CARD01060092[redacted],PR,PURCHASE *,$65.00,703340,1832703340,H[redacted]
2017-03-30,15:00:11,CARD01054749[redacted]R,PURCHASE,$5.88,005551,0000005634,300V76[redacted]
2017-03-30,15:03:19,CARD01050131[redacted]ASE,$12.66,629169,0150629169,HPSC00307000[redacted]
2017-03-30,15:03:42,CARD01059779[redacted]RCHASE,$19.68,005922,0005922251,HATHMSAL6[redacted]
2017-03-30,15:03:55,CARD01060091[redacted]UAS,PR,PURCHASE,$40.00,004470,0000004406,[redacted]
2017-03-30,15:04:45,CARD01060092[redacted]N 5495,PR,PURCHASE,$29.52,025085,00000251[redacted]
2017-03-30,15:06:45,CARD01054091[redacted]CHASE,$6.98,002073,0000002080,300V8738[redacted]
2017-03-30,15:06:49,CARD01058043[redacted]43,PR,PURCHASE,$6.67,047600,0000047857,30[redacted]
2017-03-30,15:07:09,CARD01051320[redacted]USE,PR,PURCHASE,$13.37,003596,0000003614,[redacted]
2017-03-30,15:07:34,CARD01055482[redacted]LOREN,PR,PURCHASE,$6.38,016182,000001616З[redacted]
2017-03-30,15:08:09,CARD01054619[redacted]PONCE,PR,PURCHASE,$4.25,007728,0000007948[redacted]
2017-03-30,15:09:30,CARD01060091[redacted]PR,PURCHASE,$64.04,003518,0000003699,30V2[redacted]
2017-03-30,15:10:32,CARD01060060[redacted]INC,PR,PURCHASE,$7.81,015798,0000016114,3[redacted]
2017-03-30,15:10:41,CARD01060092[redacted]LS LLC,PR,PURCHASE,$6.30,038009,000003711[redacted]
2017-03-30,15:11:20,CARD01060090[redacted]N 3774,PR,PURCHASE,$8.56,000762,000000076[redacted]
2017-03-30,15:11:28,CARD01060092[redacted]SE,PR,PURCHASE,$16.21,000852,0000000923,3[redacted]
2017-03-30,15:11:30,CARD01060091[redacted]NAS ECR,PR,PURCHASE,$59.26,009449,0330009[redacted]0009
2017-03-30,15:11:56,CARD01060060[redacted]URCHASE,$44.15,023491,0000023514,30V22933[redacted]
2017-03-30,15:11:57,CARD01060090[redacted]A,PR,PURCHASE,$6.50,036182,0000038040,30V[redacted]
2017-03-30,15:12:07,CARD01060092[redacted]S SERVIC,PR,PURCHASE,$21.66,024245,000002[redacted]
2017-03-30,15:12:11,CARD01060091[redacted],PR,PURCHASE,$7.40,004921,0330004921,HATH[redacted]
2017-03-30,15:12:39,CARD01060092[redacted]CATALINA,PR,PURCHASE,$5.24,006584,0000006[redacted]
2017-03-30,15:13:20,CARD01053311[redacted] BAY,PR,PURCHASE,$10.00,029047,0000026272[redacted]
2017-03-30,15:13:38,CARD01060060[redacted]T Y FRIGA,PR,PURCHASE,$183.70,012170,0000[redacted]
2017-03-30,15:13:46,CARD01060090[redacted]S SS,PR,PURCHASE,$15.00,264381,0000258650[redacted]
2017-03-30,15:13:47,CARD01060091[redacted]TOLIO,PR,PURCHASE,$7.63,020541,0000020651[redacted]
2017-03-30,15:13:53,CARD01054633[redacted] ECR,PR,PURCHASE,$28.67,005932,0330005932[redacted]5
```

1: Detailed Transactions Dashboard

Dashboard allows to do necessary filtering and searching for transactions data

splunk> App: Security Essentials Anti-Fraud | Administrator | 1 Messages | Settings | Activity | Help | Find

Search | Anti-Fraud Scenarios | Dashboards | Anti-Fraud Scenarios - Links | Security Essentials Anti-Fraud

Payment Cards: Detailed Transactions

Edit | Export | ...

Custom filter. Ex: *8016 OR Walmart: *

Select Card (top 250 only): Any Card

Select Merchant (top 250 only): Any Merchant

Compromised payment cards: Show everything / Show only compromised cards

Display cards by risk: Show all payment cards / Show only risky cards

Select time period (of available data): Last 2 days (full txn data)

Summarize card data: Show detailed data / Summarize all data

Limit number of results: Show last 5000 results

Hide Filters

Reset Dashboard

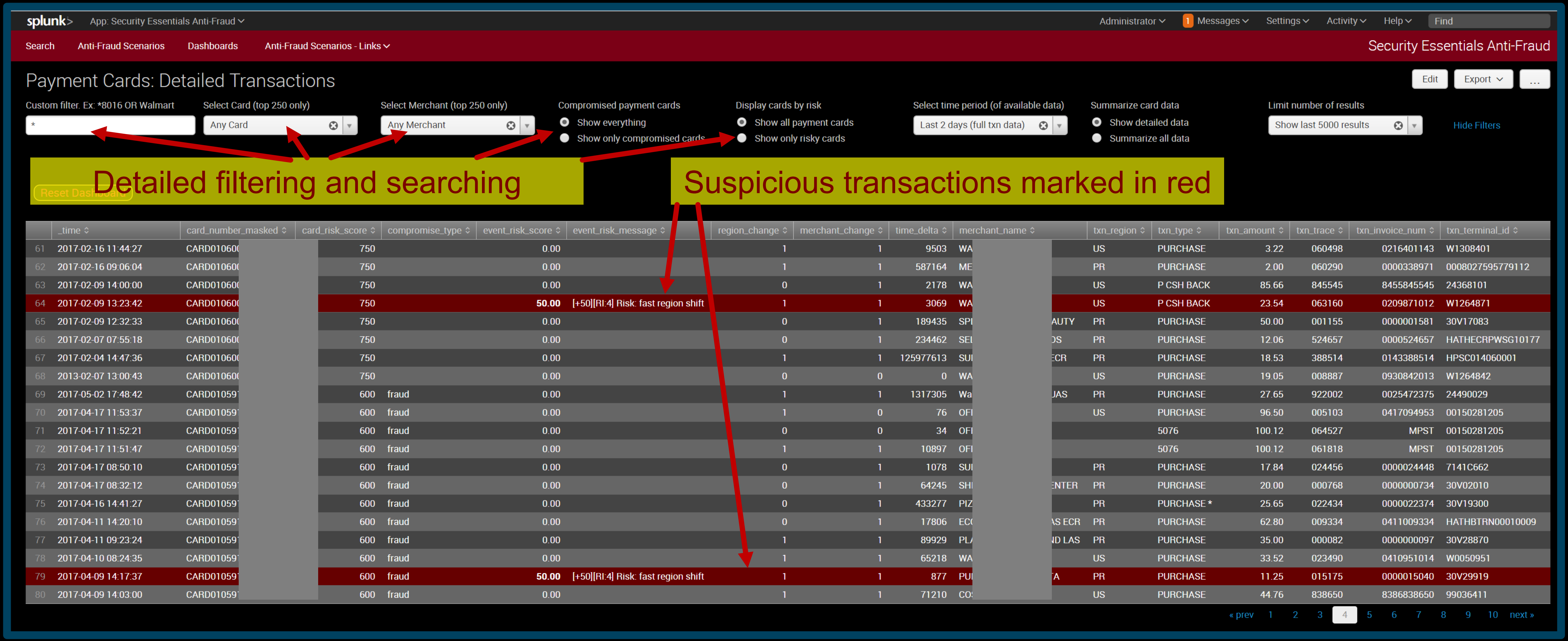
	_time	card_number_masked	card_risk_score	compromise_type	event_risk_score	event_risk_message	region_change	merchant_change	time_delta	merchant_name	txn_region	txn_type	txn_amount	txn_trace	txn_invoice_num	txn_terminal_id
61	2017-02-16 11:44:27	CARD01060	750		0.00		1	1	9503	WA	US	PURCHASE	3.22	060498	0216401143	W1308401
62	2017-02-16 09:06:04	CARD01060	750		0.00		1	1	587164	ME	PR	PURCHASE	2.00	060290	0000338971	0008027595779112
63	2017-02-09 14:00:00	CARD01060	750		0.00		0	1	2178	WA	US	P CSH BACK	85.66	845545	8455845545	24368101
64	2017-02-09 13:23:42	CARD01060	750		**50.00**	[+50][RI:4] Risk: fast region shift	1	1	3069	WA	US	P CSH BACK	23.54	063160	0209871012	W1264871
65	2017-02-09 12:32:33	CARD01060	750		0.00		0	1	189435	SPI AUTY	PR	PURCHASE	50.00	001155	0000001581	30V17083
66	2017-02-07 07:55:18	CARD01060	750		0.00		0	1	234462	SEI DS	PR	PURCHASE	12.06	524657	0000524657	HATHECRPWSG10177
67	2017-02-04 14:47:36	CARD01060	750		0.00		1	1	125977613	SUI ECR	PR	PURCHASE	18.53	388514	0143388514	HPSC014060001
68	2013-02-07 13:00:43	CARD01060	750		0.00		0	0	0	WA	US	PURCHASE	19.05	008887	0930842013	W1264842
69	2017-05-02 17:48:42	CARD01059	600	fraud	0.00		1	1	1317305	Wa UAS	PR	PURCHASE	27.65	922002	0025472375	24490029
70	2017-04-17 11:53:37	CARD01059	600	fraud	0.00		1	0	76	OFI	US	PURCHASE	96.50	005103	0417094953	00150281205
71	2017-04-17 11:52:21	CARD01059	600	fraud	0.00		0	0	34	OFI		5076	100.12	064527	MPST	00150281205
72	2017-04-17 11:51:47	CARD01059	600	fraud	0.00		1	1	10897	OFI		5076	100.12	061818	MPST	00150281205
73	2017-04-17 08:50:10	CARD01059	600	fraud	0.00		0	1	1078	SUI	PR	PURCHASE	17.84	024456	0000024448	7141C662
74	2017-04-17 08:32:12	CARD01059	600	fraud	0.00		0	1	64245	SHI ENTER	PR	PURCHASE	20.00	000768	0000000734	30V02010
75	2017-04-16 14:41:27	CARD01059	600	fraud	0.00		0	1	433277	PIZ	PR	PURCHASE *	25.65	022434	0000022374	30V19300
76	2017-04-11 14:20:10	CARD01059	600	fraud	0.00		0	1	17806	EC AS ECR	PR	PURCHASE	62.80	009334	0411009334	HATHBTRN00010009
77	2017-04-11 09:23:24	CARD01059	600	fraud	0.00		1	1	89929	PLA ND LAS	PR	PURCHASE	35.00	000082	0000000097	30V28870
78	2017-04-10 08:24:35	CARD01059	600	fraud	0.00		1	1	65218	WA	US	PURCHASE	33.52	023490	0410951014	W0050951
79	2017-04-09 14:17:37	CARD01059	600	fraud	**50.00**	[+50][RI:4] Risk: fast region shift	1	1	877	PUI A	PR	PURCHASE	11.25	015175	0000015040	30V29919
80	2017-04-09 14:03:00	CARD01059	600	fraud	0.00		1	1	71210	CO	US	PURCHASE	44.76	838650	8386838650	99036411

« prev 1 2 3 4 5 6 7 8 9 10 next »

splunk> .conf2017

1: Detailed Transactions Dashboard

Dashboard allows to do necessary filtering and searching for transactions data



	_time	card_number_masked	card_risk_score	compromise_type	event_risk_score	event_risk_message	region_change	merchant_change	time_delta	merchant_name	txn_region	txn_type	txn_amount	txn_trace	txn_invoice_num	txn_terminal_id
61	2017-02-16 11:44:27	CARD010600 [illegible]	750		0.00		1	1	9503	WA [illegible]	US	PURCHASE	3.22	060498	0216401143	W1308401
62	2017-02-16 09:06:04	CARD010600 [illegible]	750		0.00		1	1	587164	ME [illegible]	PR	PURCHASE	2.00	060290	0000338971	0008027595779112
63	2017-02-09 14:00:00	CARD010600 [illegible]	750		0.00		0	1	2178	WA [illegible]	US	P CSH BACK	85.66	845545	8455845545	24368101
64	2017-02-09 13:23:42	CARD010600 [illegible]	750		**50.00**	[+50][RI:4] Risk: fast region shift	1	1	3069	WA [illegible]	US	P CSH BACK	23.54	063160	0209871012	W1264871
65	2017-02-09 12:32:33	CARD010600 [illegible]	750		0.00		0	1	189435	SPI [illegible] AUTY	PR	PURCHASE	50.00	001155	0000001581	30V17083
66	2017-02-07 07:55:18	CARD010600 [illegible]	750		0.00		0	1	234462	SEI [illegible] OS	PR	PURCHASE	12.06	524657	0000524657	HATHECRPWSG10177
67	2017-02-04 14:47:36	CARD010600 [illegible]	750		0.00		1	1	125977613	SUI [illegible] ECR	PR	PURCHASE	18.53	388514	0143388514	HPSC014060001
68	2013-02-07 13:00:43	CARD010600 [illegible]	750		0.00		0	0	0	WA [illegible]	US	PURCHASE	19.05	008887	0930842013	W1264842
69	2017-05-02 17:48:42	CARD010591 [illegible]	600	fraud	0.00		1	1	1317305	Wa [illegible] JAS	PR	PURCHASE	27.65	922002	0025472375	24490029
70	2017-04-17 11:53:37	CARD010591 [illegible]	600	fraud	0.00		1	0	76	OFI [illegible]	US	PURCHASE	96.50	005103	0417094953	00150281205
71	2017-04-17 11:52:21	CARD010591 [illegible]	600	fraud	0.00		0	0	34	OFI [illegible]		5076	100.12	064527	MPST	00150281205
72	2017-04-17 11:51:47	CARD010591 [illegible]	600	fraud	0.00		1	1	10897	OFI [illegible]		5076	100.12	061818	MPST	00150281205
73	2017-04-17 08:50:10	CARD010591 [illegible]	600	fraud	0.00		0	1	1078	SUI [illegible]	PR	PURCHASE	17.84	024456	0000024448	7141C662
74	2017-04-17 08:32:12	CARD010591 [illegible]	600	fraud	0.00		0	1	64245	SHI [illegible] ENTER	PR	PURCHASE	20.00	000768	0000000734	30V02010
75	2017-04-16 14:41:27	CARD010591 [illegible]	600	fraud	0.00		0	1	433277	PIZ [illegible]	PR	PURCHASE *	25.65	022434	0000022374	30V19300
76	2017-04-11 14:20:10	CARD010591 [illegible]	600	fraud	0.00		0	1	17806	ECI [illegible] AS ECR	PR	PURCHASE	62.80	009334	0411009334	HATHBTRN00010009
77	2017-04-11 09:23:24	CARD010591 [illegible]	600	fraud	0.00		1	1	89929	PLA [illegible] ND LAS	PR	PURCHASE	35.00	000082	0000000097	30V28870
78	2017-04-10 08:24:35	CARD010591 [illegible]	600	fraud	0.00		1	1	65218	WA [illegible]	US	PURCHASE	33.52	023490	0410951014	W0050951
79	2017-04-09 14:17:37	CARD010591 [illegible]	600	fraud	**50.00**	[+50][RI:4] Risk: fast region shift	1	1	877	PUI [illegible] TA	PR	PURCHASE	11.25	015175	0000015040	30V29919
80	2017-04-09 14:03:00	CARD010591 [illegible]	600	fraud	0.00		1	1	71210	COS [illegible]	US	PURCHASE	44.76	838650	8386838650	99036411

« prev 1 2 3 4 5 6 7 8 9 10 next »

Detecting Suspicious Transactions With Splunk

How we can detect suspicious transactions?

1. We know approximate location of each transaction. In this case: State
2. If time between 2 adjacent transactions is too small considering their physical location – Alert!

0.00		0	1	2178	WAL-MART #3681	US	P CSH BACK
50.00	[+50][RI:4] Risk: fast region shift	1	1	3069	WALGREENS STORE	US	P CSH BACK
0.00		0	1	189435	SPLENDOR PLAZA BEAUTY	PR	PURCHASE
0.00		0	1	234462	SELECTOS LOS PRADOS	PR	PURCHASE

Order of Txns: Newest on top

Splunk detected suspicious transaction passed at a physical location within impossibly short time from transaction done in another state

Time delta between transactions

Merchant

Txn State

Txn type

How To Detect Suspicious Transactions With SPL

```
. . . . .

| sort 0 card_id, _time
| streamstats window=2 current=1
    dc(txn_region)    as region_change,
    dc(merchant_name) as merchant_change,
    range(_time)      as time_delta
  by card_id

| eval region_change=region_change-1, merchant_change=merchant_change-1

| eval risk_1_triggered=if(region_change>0   AND time_delta<7200, 1, 0)
| eval risk_2_triggered=if(merchant_change>0 AND time_delta<60,   1, 0)
```

2: Cards Risk Summary Dashboard

Allows executive to see current overall exposure to risk based on activity patters

splunk> App: Security Essentials Anti-Fraud Administrator Messages Settings Activity Help Find

Search | Anti-Fraud Scenarios | Dashboards | Anti-Fraud Scenarios - Links | Security Essentials Anti-Fraud

Payment Cards: Risk Analysis

Edit | Export | ...

Compromised payment cards: Show everything

Display cards by risk: Show only risky cards

Select time period (of available data): Last 2 days (full txn data)

Hide Filters

Reset Dashboard

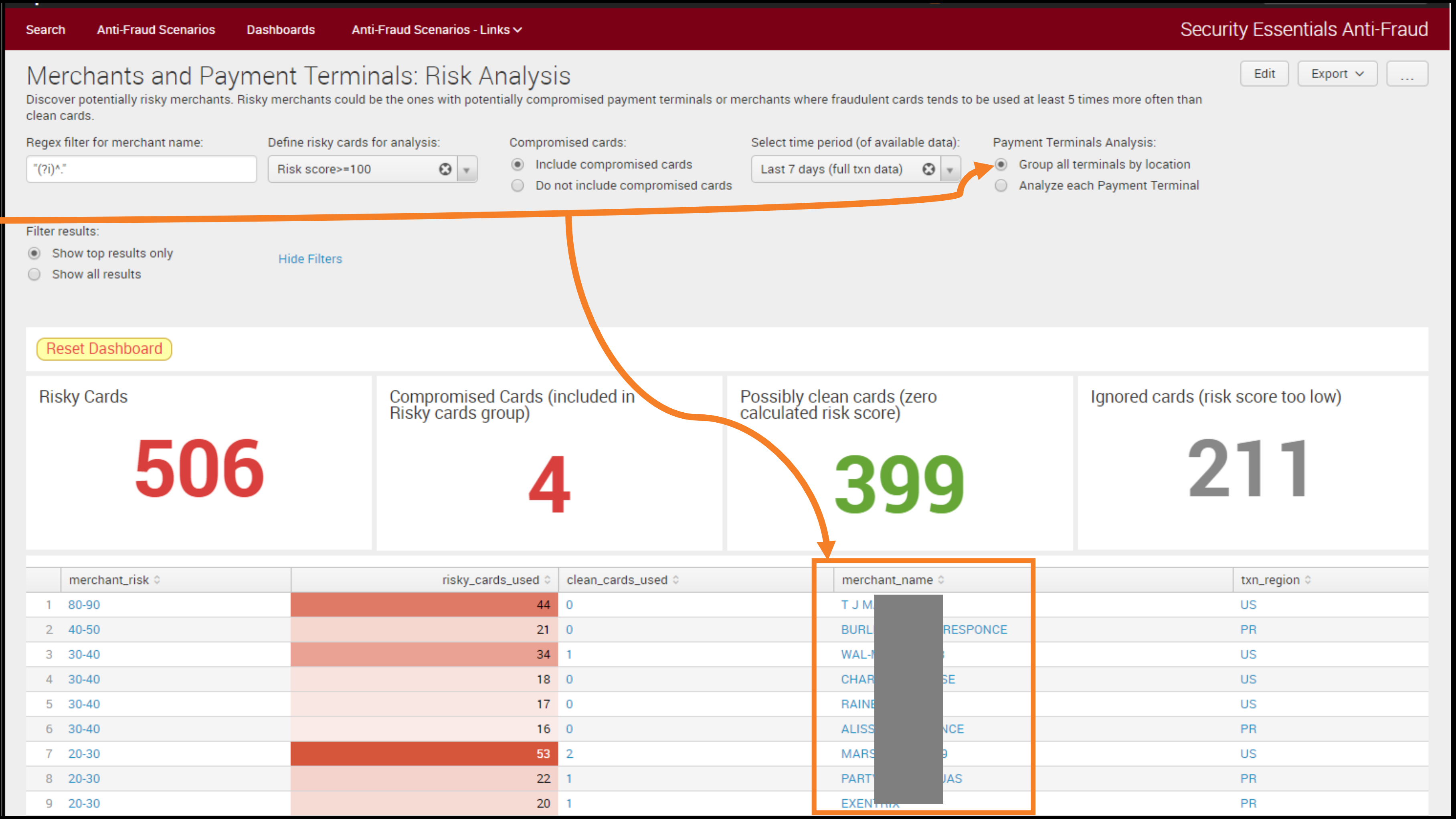
Dashboard offers single view of all risky and compromised cards sorted by risk score. Activity summary is shown for each card

	card_number_masked	card_risk_score	card_risk_messages	compromise_type	Total number of transactions	Total value of all transactions	Smallest transaction	Largest transaction	Average value of transactions
1	CARD01060[redacted]	750	15 times: [+50][RI:4] Risk: fast region shift		68	2809.51	2.00	333.20	41.32
2	CARD01059[redacted]	600	2 times: [+50][RI:4] Risk: fast region shift [+500] Marked as: fraud	fraud	60	1752.97	5.11	181.52	29.22
3	CARD01053[redacted]	500	[+500] Marked as: fraud	fraud	72	1659.84	2.53	196.70	23.05
4	CARD01060[redacted]	465	1 times: [+15][RI:5] Risk: fast merchant shift 9 times: [+50][RI:4] Risk: fast region shift		64	2048.61	3.37	128.22	32.01
5	CARD01060[redacted]	350	7 times: [+50][RI:4] Risk: fast region shift		53	1942.52	4.98	199.12	36.65
6	CARD01056[redacted]	250	5 times: [+50][RI:4] Risk: fast region shift		86	1003.60	0.33	107.69	11.67
7	CARD01060[redacted]	230	2 times: [+15][RI:5] Risk: fast merchant shift 4 times: [+50][RI:4] Risk: fast region shift		125	2877.03	1.99	131.00	23.02
8	CARD01060[redacted]	200	4 times: [+50][RI:4] Risk: fast region shift		32	630.56	0.95	62.79	19.71
9	CARD01060[redacted]	200	4 times: [+50][RI:4] Risk: fast region shift		52	971.06	2.98	51.19	18.67
10	CARD01060[redacted]	150	3 times: [+50][RI:4] Risk: fast region shift		63	1994.74	3.23	744.82	31.66
11	CARD01060[redacted]	150	3 times: [+50][RI:4] Risk: fast region shift		65	1412.12	1.52	500.00	21.72
12	CARD01060[redacted]	150	3 times: [+50][RI:4] Risk: fast region shift		45	1007.63	2.90	103.22	22.39
13	CARD01052[redacted]	100	2 times: [+50][RI:4] Risk: fast region shift		34	989.30	3.33	149.84	29.10
14	CARD01060[redacted]	100	2 times: [+50][RI:4] Risk: fast region shift		35	876.94	4.00	139.80	25.06
15	CARD01052[redacted]	100	2 times: [+50][RI:4] Risk: fast region shift		46	1429.53	2.43	391.20	31.08
16	CARD01060[redacted]	50	1 times: [+50][RI:4] Risk: fast region shift		82	1655.26	2.01	170.00	20.19
17	CARD01052[redacted]	50	1 times: [+50][RI:4] Risk: fast region shift		27	855.01	5.84	105.90	31.67

3: Merchants and Payment Terminals Analysis

Detect anomalies of card usage at specific merchants and payment terminals

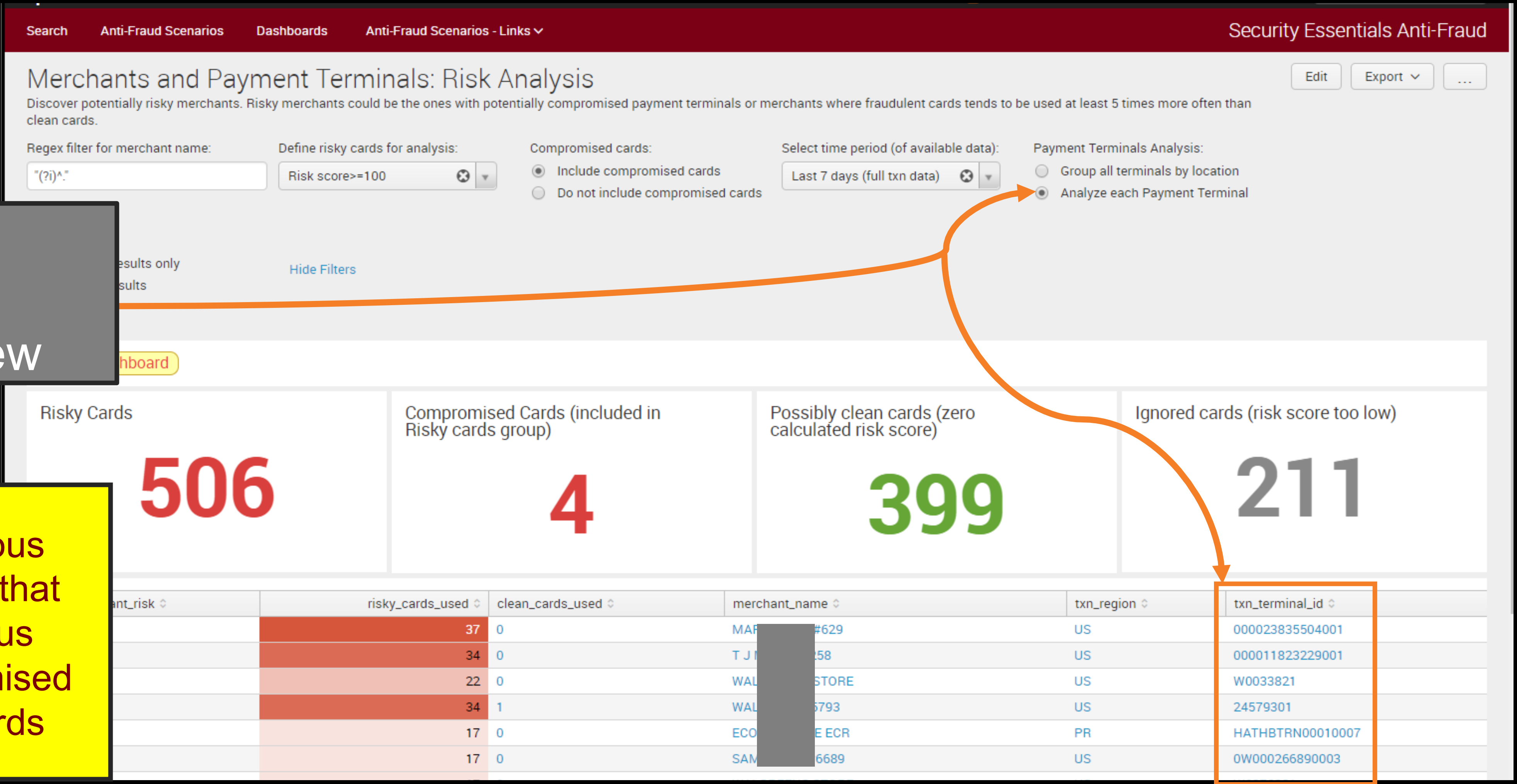
Suspicious Merchants View



Risk scoring of merchants and payment terminals that process excessive amounts of compromised and risky behaving payment cards

3: Merchants and Payment Terminals Analysis, cont.

Detect anomalies of card usage at specific merchants and payment terminals



Suspicious Payment Terminals View

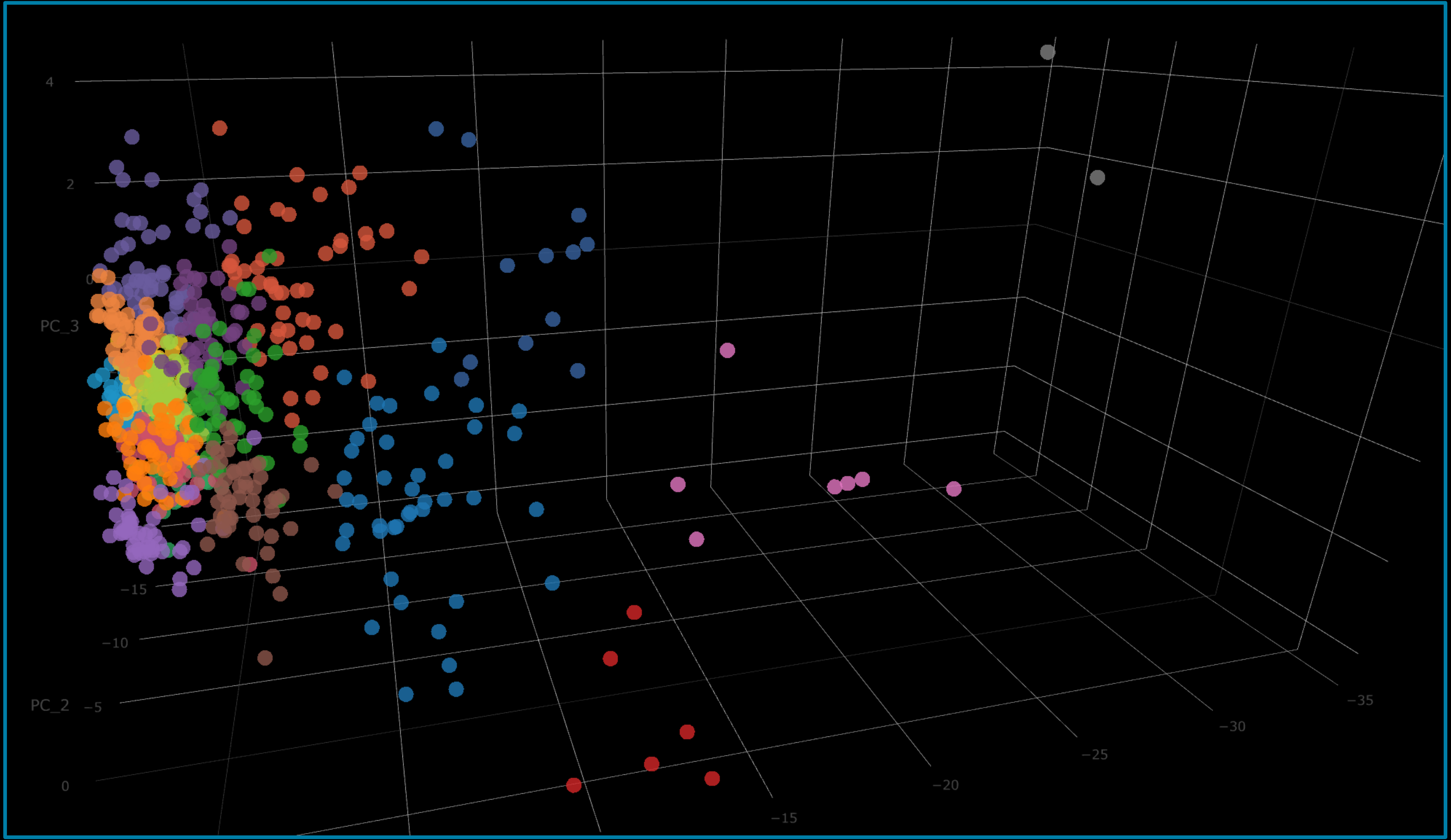
Analyzing suspicious payment terminals that process anomalous number of compromised cards vs. other cards

4. Detecting Anomalous Behaviors

Applying unsupervised learning techniques to detect anomalous behavior and new, previously unknown fraud patterns

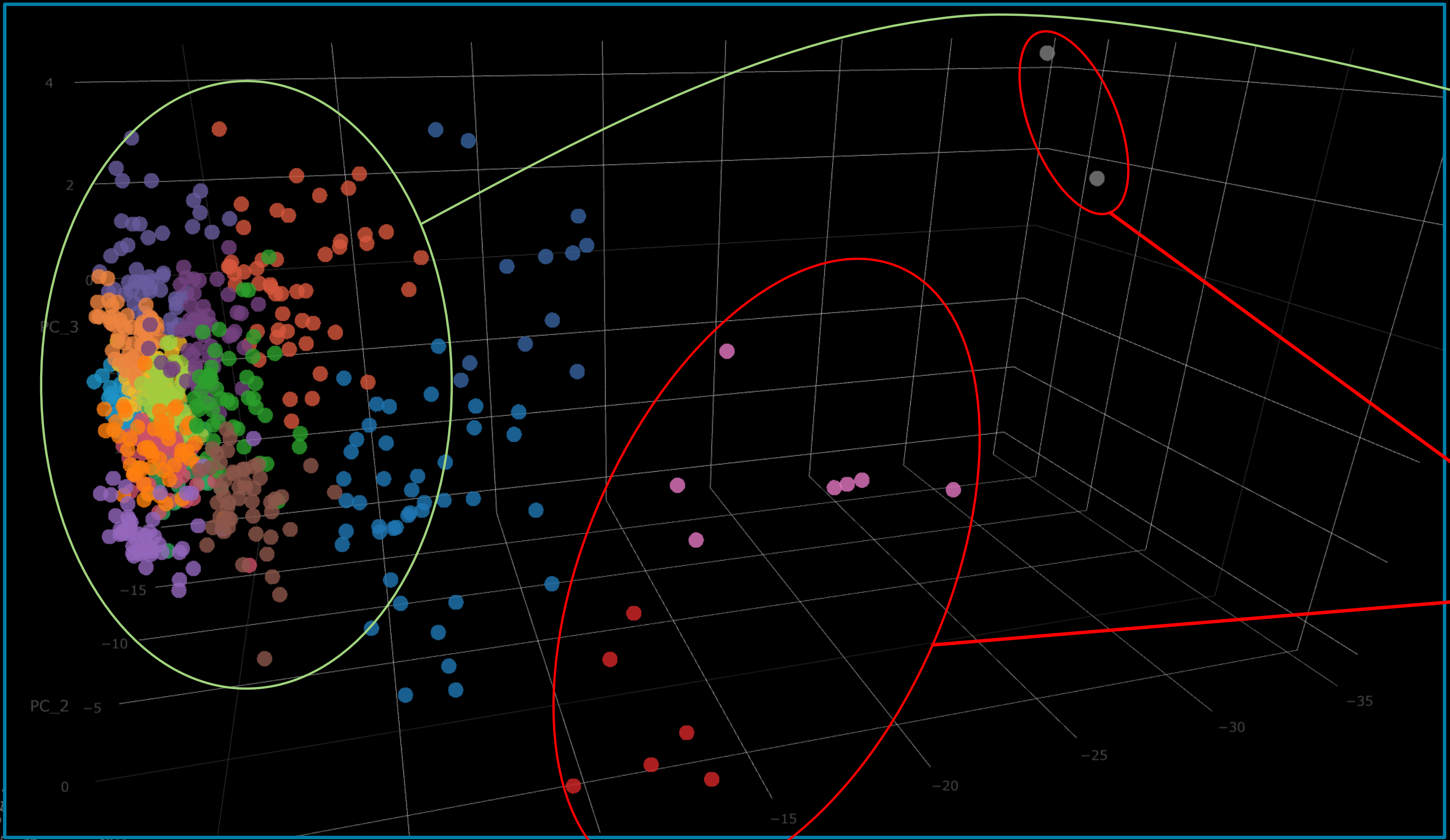
We want to be able to aggregate multidimensional behavior of all payment cards together to discover unusual, potentially risky or fraudulent behavior.

We need to simultaneously analyze multiple characteristics of all cards and all transactions and all behaviors to detect outliers and prevent potential losses.



4. Detecting Anomalous Behaviors, Cont.

Applying unsupervised learning techniques to detect anomalous behavior and new, unknown fraud patterns



“Normal” or typical behaviors are grouped together

Anomalous behaviors stands out from the majority of the crowd.

Detecting Anomalies Via Clustering

Applying Machine Learning Toolkit clustering to filter our anomalies

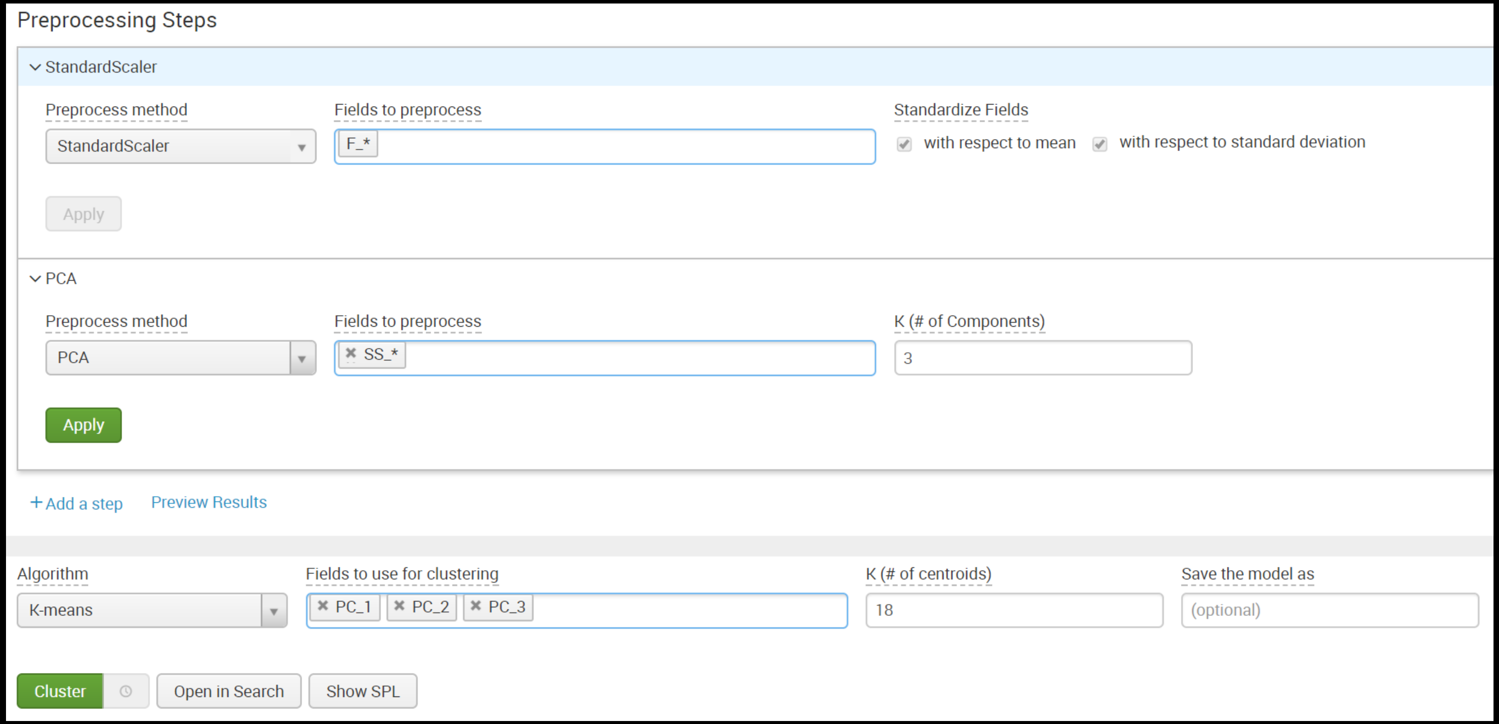
```
index=vpnet2 sourcetype=cards_txn2
| where len(txn_region)>0 | dedup _raw | sort 0 card_id, _time
| streamstats
  window=2 current=1 dc(txn_region) as region_change,
  dc(merchant_name) as merchant_change, range(_time) as time_delta by card_id

| eval region_change=region_change-1, merchant_change=merchant_change-1

| where time_delta>0 | eval x="Throw away oldest event for each card“
| stats  c as num_txns
  max(txn_amount) as F_txn_amt_max, avg(txn_amount) as F_txn_amt_avg, stdev(txn_amount) as N_txn_amt_std
  median(txn_amount) as F_txn_amt_median,  avg(time_delta) as N_td_avg, stdev(time_delta) as N_td_std
  c(eval(merchant_change>0)) as merchant_changes_num  c(eval(region_change>0)) as region_changes_num
  by card_id

| where num_txns>=5

| eval F_merchant_changes_num_norm = merchant_changes_num / num_txns
| eval F_region_changes_num_norm   = region_changes_num / num_txns
| eval F_txn_amt_std_norm          = N_txn_amt_std / F_txn_amt_avg
| eval F_time_diff_std_norm        = N_td_std / N_td_avg
```

1

Get data!
Create SPL search
Extract needed **features**

Detecting Anomalies Via Clustering, Cont.

Applying Machine Learning Toolkit clustering to filter our anomalies



2

Apply preprocessing steps to normalize features
Define clustering algorithm

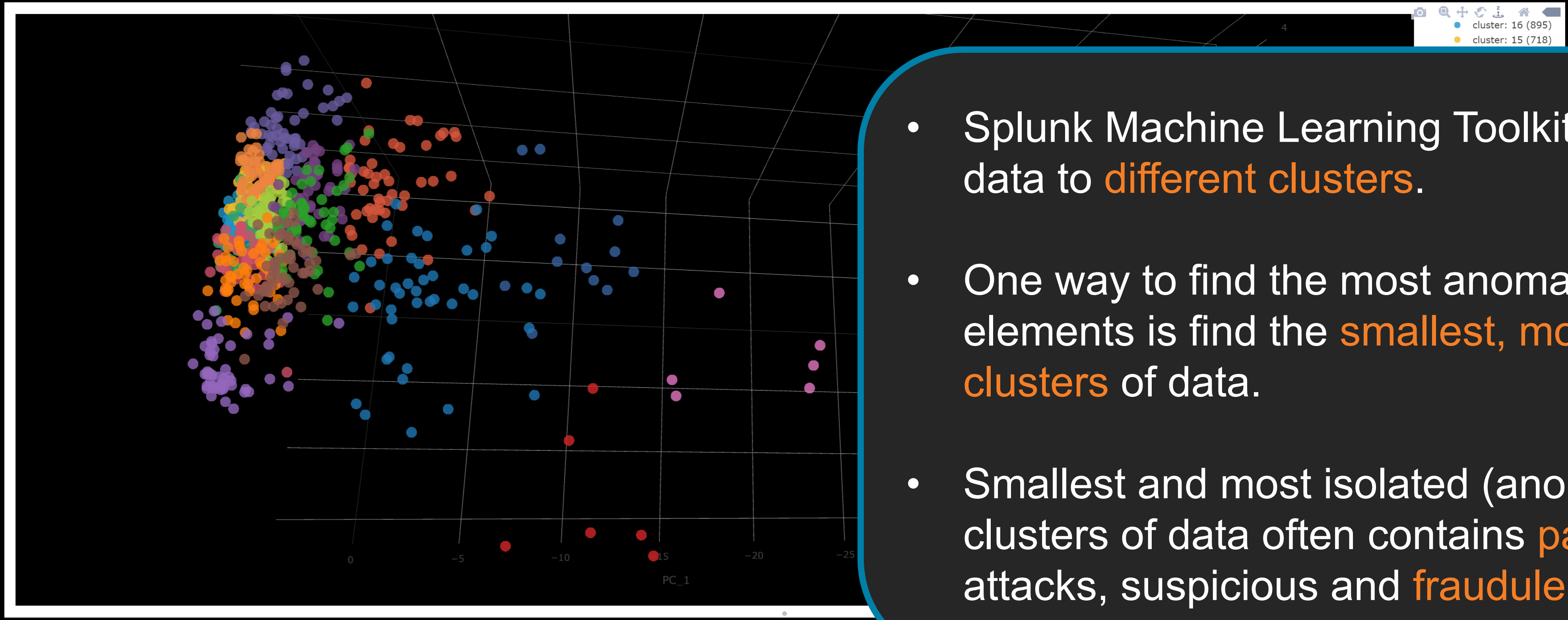
```
| fit StandardScaler F_*
| fit PCA SS_* k=3
| fit KMeans PC_1, PC_2, PC_3 k=18
```

Detecting Anomalies Via Clustering, Cont.

Applying Machine Learning Toolkit clustering to filter our anomalies

3 Get generated SPL code

Detecting Anomalies Via Clusteing, Cont.

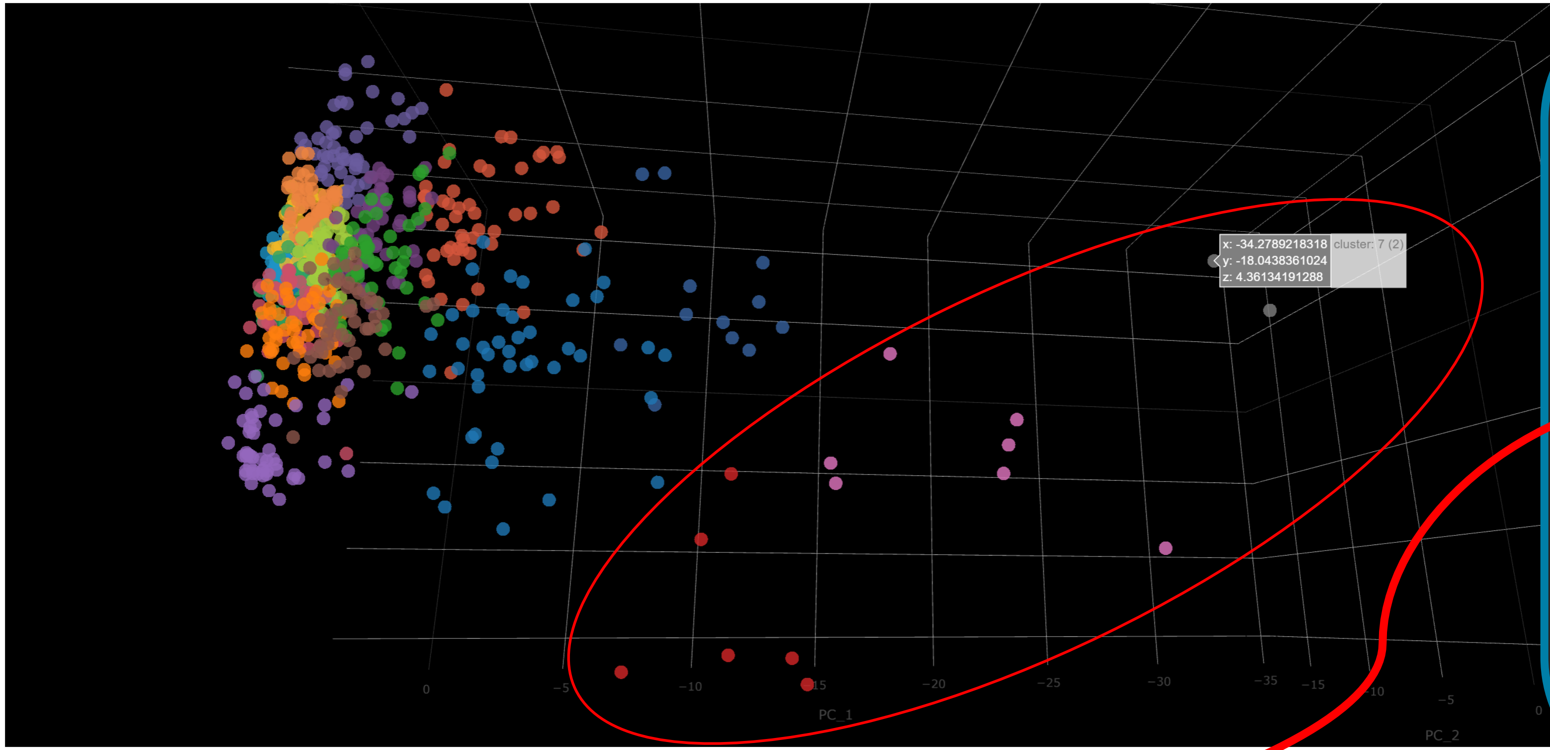


- Splunk Machine Learning Toolkit assigns data to different clusters.
- One way to find the most anomalous data elements is find the smallest, most isolated clusters of data.
- Smallest and most isolated (anomalous) clusters of data often contains patterns of attacks, suspicious and fraudulent activity.

Detected Anomalies:

	card_id	num_txns	txn_max_avg	cluster	PC_1	PC_2	PC_3	fast_merchant_change_num	fast_region_change_num
1	20862038	17	4000.0 / 926.32	7 (2)	-34.2789218318	-18.0438361024	4.36134191288	0	0
2	59076083	43	5673.36 / 1018.72	7 (2)	-33.5333291418	-10.9183802133	1.77635274178	0	2
3	56487074	149	5067.1 / 75.11	17 (6)	-13.3710615373	6.91987112988	-6.34358538491	0	9
4	17328142	140	4400.0 / 68.68	17 (6)	-11.6935752193	6.42658452449	-6.0039832731	0	5
5	7282013	69	5000.0 / 126.76	17 (6)	-13.1182666547	4.83803888853	-6.07122931579	1	4
6	562098	49	3232.33 / 83.98	17 (6)	-9.49444771347	6.28905500335	-6.19625556637	0	0
7	71819046	24	3841.4 / 183.41	17 (6)	-11.5595455213	4.56071529506	-3.61648920208	1	4
8	50013054	38	3795.36 / 132.95	17 (6)	-10.9249645555	4.28231142744	-4.51249033806	0	3
9	7540021	5	5000.0 / 1101.79	4 (7)	-25.6723393798	-4.44819051697	-4.25534255648	0	0

Detecting Anomalies Via Clustering, Cont.

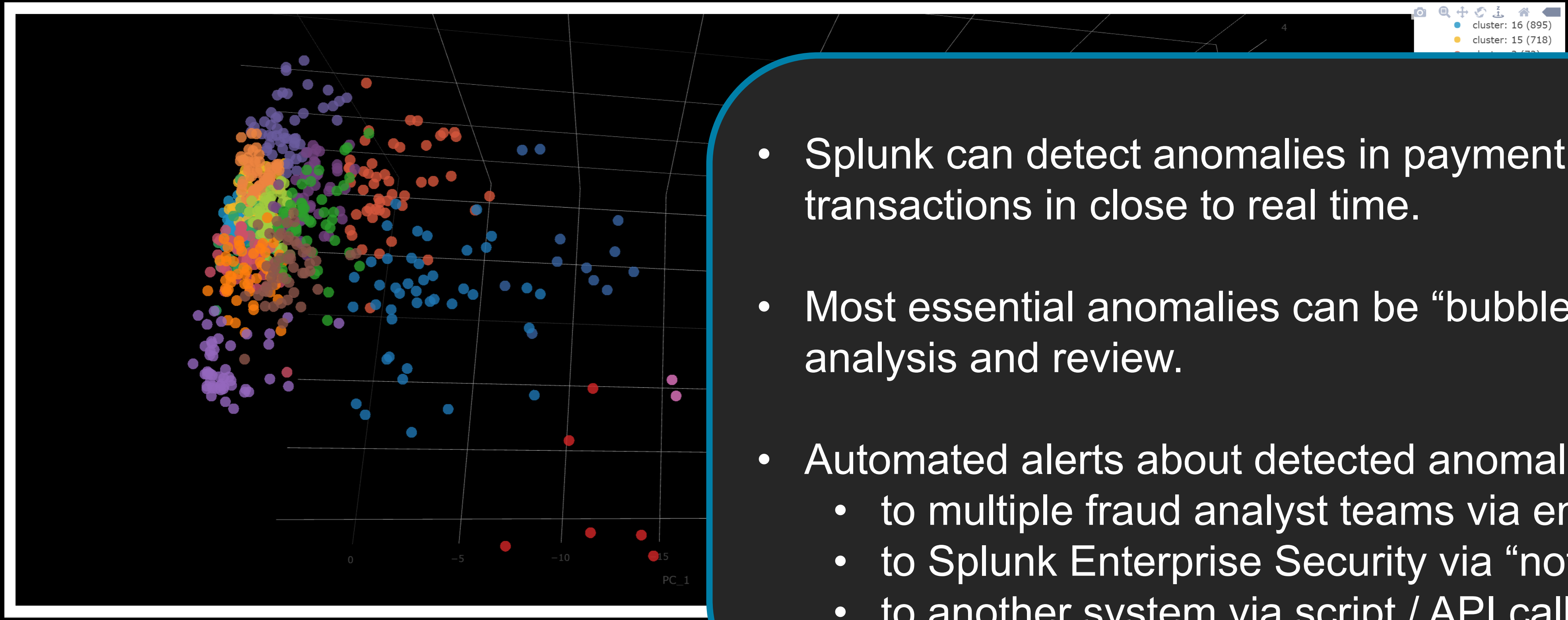


- Smallest clusters in this data representing real world dataset of credit card transactions containing patterns of suspicious activity.
- Anomalous clusters immediately shows
  - Cards with unusually high transactions values
  - Cards containing "fast region shift" fraud pattern.
  - Cards with unusual geo travel patterns
- No pre-programmed rules being used.

Detected Anomalies:

	card_id	num_txns	txn_max_avg	cluster	PC_1	PC_2	PC_3	fast_merchant_change_num	fast_region_change_num
1	20862038	17	4000.0 / 926.32	7 (2)	-34.2789218318	-18.0438361024	4.36134191288	0	0
2	59076083	43	5673.36 / 1018.72	7 (2)	[illegible]	-10.9183802133	1.77635274178	0	2
3	56487074	149	5067.1 / 75.11	[illegible]	-13.3710615373	6.91987112988	-6.34358538491	0	9
4	17328142	140	4400.0 / 68.68	17 (6)	-11.6935752193	6.42658452449	-6.0039832731	0	5
5	7282013	69	5000.0 / 126.76	17 (6)	-13.1182666547	4.83803888853	-6.07122931579	1	4
6	562098	49	3232.33 / 83.98	17 (6)	-9.49444771347	6.28905500335	-6.19625556637	0	0
7	71819046	24	3841.4 / 183.41	17 (6)	-11.5595455213	4.56071529506	-3.61648920208	1	4
8	50013054	38	3795.36 / 132.95	17 (6)	-10.9249645555	4.28231142744	-4.51249033806	0	3
9	7540021	5	5000.0 / 1101.79	4 (7)	-25.6723393798	-4.44819051697	-4.25534255648	0	0

Detecting Anomalies Via Clustering, Cont.



- Splunk can detect anomalies in payment card transactions in close to real time.
- Most essential anomalies can be “bubbled up” for analysis and review.
- Automated alerts about detected anomalies can be sent:
  - to multiple fraud analyst teams via email alerts.
  - to Splunk Enterprise Security via “notable events”
  - to another system via script / API calls.

Detected Anomalies:

	card_id	num_txns	txn_max_avg	cluster	PC_1				
1	20862038	17	4000.0 / 926.32	7 (2)	-34.2789218318				
2	59076083	43	5673.36 / 1018.72	7 (2)	-33.5333291418				
3	56487074	149	5067.1 / 75.11	17 (6)	-13.3710615373	6.91987112988	-6.34358538491	0	9
4	17328142	140	4400.0 / 68.68	17 (6)	-11.6935752193	6.42658452449	-6.0039832731	0	5
5	7282013	69	5000.0 / 126.76	17 (6)	-13.1182666547	4.83803888853	-6.07122931579	1	4
6	562098	49	3232.33 / 83.98	17 (6)	-9.49444771347	6.28905500335	-6.19625556637	0	0
7	71819046	24	3841.4 / 183.41	17 (6)	-11.5595455213	4.56071529506	-3.61648920208	1	4
8	50013054	38	3795.36 / 132.95	17 (6)	-10.9249645555	4.28231142744	-4.51249033806	0	3
9	7540021	5	5000.0 / 1101.79	4 (7)	-25.6723393798	-4.44819051697	-4.25534255648	0	0

Summary Notes And Conclusions

- Above fully custom fraud detection app:
  - Built with Splunk Enterprise
  - No coding, only Simple XML was used
  - No coding, everything was done via Web interface
  - Was built by 1 person
  - Was built in 7 days time
- Splunk Machine Learning Toolkit allows to apply both supervised and unsupervised learning techniques on top of any data.
- Payment cards fraud and any kind of suspicious activity can be predicted
- Known and Unknown Fraud = always anomaly. The secret of detecting known and unknown fraudulent patterns is to:
  - Have access to as much data as possible
  - Extract relevant features of behavior
  - Apply and combine anomaly detection techniques available on top of data
- Splunk Machine Learning allows to learn from data and generalize from complex data examples to predict outcomes (such as fraud, chargebacks, etc…)

Splunk Enterprise allows building of advanced, fully customized security and anti-fraud solutions in a short period of time.

splunk> .conf2017

Detecting Credit Card Fraud on Credit Unions

Preventing fraud by analyzing data on Splunk with Indicators of compromise

Felipe J . Hernandez , CEO, VPNet Inc

VPNet

SECURING EVERY NETWORK

VPNet offers innovative solutions in the telecom and IT security industry.

Serving customers across all the industries from Retail to banking.

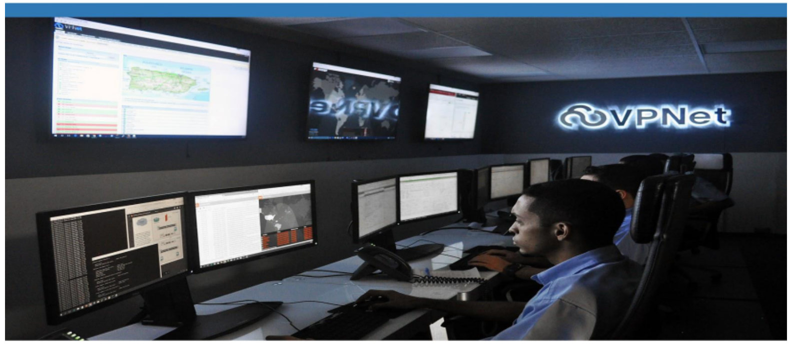
- Leaders in Cybersecurity for retail customers
- Serving 65% of the credit unions in Puerto Rico
- Leaders in Cybersecurity for the Healthcare market

Our success and having contracts with reputable companies in Puerto Rico, it is directly related to our commitment with the quality and excellent customer service.

WWW.VPNET.NET

VPNet

THE POWER OF THE CLOUD



SECURITY OPERATIONS CENTER:

VPNet entiende la importancia de mantener los sistemas y redes de nuestros clientes con los mejores estándares de seguridad. Con el Security Operations Center, VPNet asegura que su red esté monitoreada en todo momento. El SOC se ocupa de mantener bajo vigilancia constante los diferentes activos de la cooperativa al reducir el tiempo de respuesta a los ataques, minimizando así las consecuencias de los mismos.

- Monitoreo de la red 24/7
- Control de cambios
- IPS - Sistema de Prevención de Intrusos
- Monitoreo de terminales
- IDS - Sistema de Detección de Intrusos
- Monitoreo de móviles
- Cumplimiento con PCI
- Reportes de su red

Teléfono: 787-620-5950
Email: info@vpnet.net

Debit Card Fraud in Credit Unions

The Problem: MasterCard Brand debit cards suffering from massive fraud issues

Impact:

- Debit card losses not protected by MC Credit insurance
- Losses not covered by local clearing house
- Credit union covering 100% of the losses
- Big impact on CU image plus inconveniences for customers
- High cost of replacement of compromised cards, average of $35 per replacement

Only Limited Tools Available To Protect Credit Unions

Technical tools:

- Using technical tools to stop transaction based on human suspicion:
  - Falcon
  - TEXT message
  - MC interface to block countries and vendors
  - Limited spending control

Business intelligence tools:

- Transactional Data history N/A
  - No average transaction amount, 20pt
  - No demographic info
  - NO spending patterns
  - NO risk assessment of customers

How We Could Help Credit Unions?

Creating a tool that would help them minimize their risk that would:

- Provide historical Data on users
- Spending patterns
- Other IOC’s that could create a riskier profile
- Using Machine learning creating a self adjusting credit risk based on behavior
- Locate which stolen cards have not been identified yet as compromised!!!

How We Could Help Credit Unions, Cont.

We need to consider:

- New compromised cards would pop up every day.
- The exposure was totaling near $300k .
- Still no clear idea on how the breach happened.
- Since all cards couldn't be voided simultaneously a maximum expending allowance was needed for users that were not classified as compromised yet, but where in risk. This allotment was going to be based on their risk score.

Getting The Data

- Live data wasn't useful without a historical perspective
- All historical data was provided in archived and proprietary form, so significant reformatting had to be done.
- We needed Gleb urgently!!
- Once historical data was inserted into Splunk, we started seeing patterns that were very insightful
- From now on Splunk stream will provide access to live data.

Crunch time!!!

What The Numbers Told Us?

- Many cards that were not reported as compromised by users
- Patterns used by fraudsters to test the cards and not alert the owner
- Merchants that were being used for the transactions
- POS used for the purpose
- POS where cleared out as the source of compromise.
- At that point data was either extracted from the institution or leaked at the clearing house level.
- Later after receiving data from other institutions that had some fraud as well, the same POS’s were also used with cards of other institutions, proving that the problem wasn't one of breaching at the CU.
- Many cards that were not reported as compromised by users

Patterns Of Fraudsters

- Tested the cards 90 days before the massive charges.
- Tested with small Amounts, 2-5 Dollars in average.
- Tested on common merchants like Walgreens, Costco and Target
- They used the same POS's (possibility of complicit behavior with cashiers).
- Use the cards to purchase Gift Cards and other debit cards to sanitize the money.
- As soon as they noticed that some cards quit working they stopped transacting and waited a few days before continuing their use.
- Typically the transaction average was $500 but the amount would vary dependent on their sense of being traced.

Splunk Benefits For VPNet

- One platform for all of our security elements
- Single point to manage all of our data intake
- One platform to manage and measure all data, from security events to interactions on our Social-Wifi Network.
- Opportunity to monetize our Data.
- Create Business intelligence solutions for customers
- Move all critical elements of our operations into Splunk

What's Next?

- Integrate Splunk STREAM as the basis of data collection for all customers
- Move the monitoring to our SOC
- Use machine learning for risk scoring to achieve automatic transaction blocking or establish funds limits.
- Create a full suite of business intelligence products.
- Integrating Splunk on Credit Unions for all their statistical analysis including marketing and members behavior.
- Expand services to all Caribbean countries

Happy Splunking!

Thank You

Don't forget to rate this session in the .conf2017 mobile app

splunk> .conf2017