Into the Deep Web: Understanding E-commerce Fraud from Autonomous Chat with Cybercriminals

Peng Wang, Xiaojing Liao, Yue Qin, XiaoFeng Wang
Indiana University Bloomington
E-commerce fraud

online fraudsters
Crowdsourcing in e-commerce fraud
Crowdsourcing via IM

Crowdsourcing via Instant Messaging (IM)
Bonus hunting

E-commerce platforms

Bonus hunters

Small-time workers

Amazon

JD.COM

eBay
Fraud account trading

Account merchants -> $\$\$\$\$\$

E-commerce platforms

Account trading storefronts

Type1 ---- $0.5
Type2 ---- $0.8
Type3 ---- $1.5
Type4 ---- $4.5

Small-time workers

Fraud account trading
SIM farming

SIM Sources:
- VoIP cards
- ...

Carriers → SIM farmers → SIM farms (websites or software) → Account merchants

$\text{SIM}$
E-commerce fraud ecosystem

- SIM farmers
  - $\rightarrow$ Small-time workers
  - $\rightarrow$ Fake transaction operators
- Account fraudsters
  - $\rightarrow$ Small-time workers
  - $\rightarrow$ E-commerce platforms

E-commerce platforms:
- Amazon
- JD.COM
- eBay
E-commerce fraud groups

Fake review groups on Telegram

Fraud account groups on QQ
E-commerce fraud group chat

Telegram

[st0rex.com] Amazon Accounts
606 subscribers

[st0rex.com] Amazon Accounts
10:33 AM

Hurry up, before sold out !!! 🔥🔥🔥

 sple BEST ACCOUNTS WITH FULL ACCESS + COOKIES sple

[st0rex.com] Amazon Accounts
12:56 PM

Few accounts added, fresh as fuck.

New price - 4$ 🎁

[st0rex.com] Amazon Accounts
1:00 PM

!! REMEMBER !! 🔥

1- All amazon accounts have active credit card attached!
2- All accounts have full access, cookie + email-access, cc expire date and everything

Group chat
Threat intelligence gathering:
collecting **evidence-based** threat information
about an existing or emerging threat

SIM farmers:
1) SIM card source
2) Gateway link/tool
3) Account merchants
4) Hack tools

Fraud account merchants:
1) Account types
2) Store link
3) Payment method
4) SIM card source
5) Hack tools
6) Fraud order tasks

Fraud account operators:
1) Fraud order tasks
2) Shipping address
3) Report link
4) Hack tools
5) Account merchants
Group chat V.S. individual chat

Telegram

[st0rex.com] Amazon Accounts
Hurry up, before sold out !!! 🔥🔥🔥

[st0rex.com] Amazon Accounts
Few accounts added, fresh as fuck.
New price - 4$ 😁

[st0rex.com] Amazon Accounts
!! REMEMBER !! 🔥
1- All amazon accounts have active credit card attached!
2- All accounts have full access, cookie + email-access, cc expire date and everything

I need XX account, what types are your selling

normal and new account
can be used for collecting coupons
and order scalp

How to buy? any self-service website

http://xxx.fakeaccountswebsites.xx

Which Sim farm did you use for registering the accounts?

private SIM farm

Ok
Do you know any tool for collecting coupons and placing orders?

XX tool name

Ok, thanks
no problem

Group chat V.S. Individual chat

account type
account store link
SIM card source
hack tool name
Intelligence gathering challenges

• **Active** intelligence gathering
  • useful intelligence is only shared through one-on-one conversation
  • the number of new fraudsters keep growing
Intelligence gathering challenges

• **Active** intelligence gathering
  • useful intelligence is only shared through one-on-one conversation
  • the number of new fraudsters keep growing

• **Automated conversation with fraudsters**
  • existing chatbots can not collect e-commerce threat intelligence
  • how to strategically lead the fraudsters to discuss the target threat intelligence is complicated
Aubrey

**Autonomous chatbot for intelligence discovery**

- **first** autonomous conversation system for active threat intel. gathering from e-commerce miscreants
- **effectively** extract great number of valuable fraud-related artifacts
- **new insights** into the e-commerce fraud ecosystem
Information exchange

SIM farmers $$ \rightarrow $$ Small-time workers $$ \rightarrow $$ Fake transaction operators $$ \rightarrow $$ E-commerce platforms

Account fraudsters $$ \rightarrow $$ Small-time workers $$ \rightarrow $$ Fake transaction operators $$ \rightarrow $$ E-commerce platforms

Amazon, eBay, JD.COM
Observation

E-commerce fraudster

Hi there
I need XX account, what types are you selling

normal and new account
can be used for collecting coupons
and order scalping

How to buy? any self-service website

http://xxx.fakeaccountswebsites.xx

Small-time worker

Which Sim farm did you use for registering the accounts?

private SIM farm

Ok

Do you know any tool for collecting coupons and placing orders?

XX tool name

Ok, thanks

no problem
Observation

Question
Hi there
I need XX account, what types are you selling
normal and new account
can be used for collecting coupons
and order scalping

Answer
How to buy? any self-service website
http://xxx.fakeaccountswebsites.xx

Question
Which Sim farm did you use for registering the accounts?
private SIM farm

Answer
Ok

Question
Do you know any tool for collecting coupons and placing orders?
XX tool name

Answer
Ok, thanks
no problem

E-commerce fraudster

Small-time worker
Architecture

Target Finder
- Feature Extraction
- Role Identification

Target Roles
- Target info
- questions
- responses

Dialog Manager
- Response Analysis
  - State Transition
  - Intel. Extraction
- Cyberthreat Intelligence

Strategy Generator
- FSM Generation
- Knowledge Source Extension
- Dialog Corpus
- Strategy
Target Finder

- Feature Extraction
- Role Identification

150 fraud IM groups

Keyword features
Behavioral features
Intent indicators

Dialog Corpus

FSM Generation
Knowledge Source Extension

Strategy Generator

Target Finder

Target Roles

Response Analysis
- State Transition
- Intel. Extraction

Dialog Manager

Questions
Responses

Target info

Cyberthreat Intelligence

February 26, 2020
Strategy Generator

Target Finder

- Feature Extraction
- Role Identification

Target Roles

- Target info
- questions
- responses

Target Finder

FSM Generation

Knowlege Source Extension

Strategy Generator

Dialog Manager

- Response Analysis
- State Transition
- Intel. Extraction

Cyberthreat Intelligence

Seed conversations
IM group chats
E-comm forum posts

Dialog Corpus

IM Msgs

target info
FSM definition

5-tuple: \((S, R, \delta, s_0, E)\)

\(S\): set of states, question Aubrey can send to the target roles

\(R\): set of responses from the target roles

\(\delta\): \(S \times R \rightarrow S\), state transition function, decide the next state

\(s_0\): start state

\(E\): end state
Hi there

I need XX account, what types are you selling

normal and new account
can be used for collecting coupons
and order scalping

How to buy? any self-service website

http://xxx.fakeaccountswebsites.xx

Which Sim farm did you use for registering the accounts?

private SIM farm

Ok

Do you know any tool for collecting coupons and placing orders?

XX tool name

Ok, thanks

no problem
Segmentation

Seed conversation  +  text clustering
Topic detection

Seed conversation + text clustering

dialog blocks

account types

storelink

Topic identification +

Hi there

I need XX account, what types are you selling

normal and new account

can be used for collecting coupons

and order scalping

SimSource

Cross-role

Which Sim farm did you use for registering the accounts?

private SIM farm

Ok

Do you know any tool for collecting coupons and placing orders?

XX tool name

Ok, thanks

no problem

http://xxx.fakeaccountswebsites.xx
Dialog Manager

Target Finder
- Feature Extraction
- Role Identification

Target Roles
- Target Roles
- Target Info
- Questions
- Responses

Target Finder
- IM Msgs
- Dialog Corpus

Strategy Generator
- FSM Generation
- Knowledge Source Extension

Response Analysis
- State Transition
- Intel. Extraction

Dialog Manager
- Cyberthreat Intelligence
Retrieval model

- FSM for retrieval model

Current state \( \times \) Response is interrogative

\[ \rightarrow \text{Retrieval model state} \]
# Evaluation

470 miscreants, 7,250 communication messages

<table>
<thead>
<tr>
<th>Category</th>
<th># miscreants</th>
<th># interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM farmers</td>
<td>185</td>
<td>2,900</td>
</tr>
<tr>
<td>Account merchants</td>
<td>130</td>
<td>2,350</td>
</tr>
<tr>
<td>Fraud order operators</td>
<td>155</td>
<td>2,000</td>
</tr>
</tbody>
</table>
## Threat intelligence analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Obtained intelligence</th>
<th>Extended intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM farmers</td>
<td>40 SIM gateways, 36 payment intel. 16 SIM card sources and inventory intel. 8 fraud</td>
<td>323K fake phone numbers</td>
</tr>
<tr>
<td></td>
<td>account websites</td>
<td>15 fraud account types, 8 payment intel.</td>
</tr>
<tr>
<td></td>
<td>1 bonus hunting automated tool</td>
<td></td>
</tr>
<tr>
<td>Account merchants</td>
<td>38 account trading websites 25 types of fraud accounts 26 payment intelligence, 10 SIM</td>
<td>150 fraud accounts, 6 hosting platforms 14K fake phone numbers</td>
</tr>
<tr>
<td></td>
<td>gateways</td>
<td>10 private APIs</td>
</tr>
<tr>
<td>Fraud order operators</td>
<td>65 targeted items 184 fraud order addresses 71 fraud order report links 4 fraud account</td>
<td>65 fraud order affiliates 8 fraud address patterns 5 hosting</td>
</tr>
<tr>
<td></td>
<td>websites</td>
<td>platforms 8 fraud account types, 4 payment intel. 10 private</td>
</tr>
<tr>
<td></td>
<td>6 bonus hunting automated tools (same as above)</td>
<td>APIs (same as above)</td>
</tr>
</tbody>
</table>

E-commerce miscreants and corresponding threat intelligence
# Intelligence from SIM farmers

<table>
<thead>
<tr>
<th>Obtained intelligence</th>
<th>Extended intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 SIM gateways, 36 payment intel.</td>
<td>323K fake phone numbers</td>
</tr>
<tr>
<td>16 SIM card sources and inventory intel.</td>
<td></td>
</tr>
<tr>
<td>8 fraud account websites</td>
<td>15 fraud account types, 8 payment intel.</td>
</tr>
<tr>
<td>1 bonus hunting automated tool</td>
<td></td>
</tr>
</tbody>
</table>

- 90% were used for account registration
- 72% accounts were used to order online
Intelligence from Account merchants

<table>
<thead>
<tr>
<th>Obtained intelligence</th>
<th>Extended intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>38 account trading websites</td>
<td>150 fraud accounts, 6 hosting platforms</td>
</tr>
<tr>
<td>25 types of fraud accounts</td>
<td></td>
</tr>
<tr>
<td>26 payment intelligence, 10 SIM gateways</td>
<td>14K fake phone numbers</td>
</tr>
<tr>
<td>5 bonus hunting automated tools</td>
<td>10 private APIs</td>
</tr>
</tbody>
</table>

Abused private APIs and hack tools never been known before
## Intelligence from Fraud operators

<table>
<thead>
<tr>
<th>Obtained intelligence</th>
<th>Extended intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>65 targeted items</td>
<td>65 fraud order affiliates</td>
</tr>
<tr>
<td>184 fraud order addresses</td>
<td>8 fraud address patterns</td>
</tr>
<tr>
<td>71 fraud order report links</td>
<td>5 hosting platforms</td>
</tr>
<tr>
<td>4 fraud account websites</td>
<td>8 fraud account types, 4 payment intel.</td>
</tr>
<tr>
<td>6 bonus hunting automated tools (same as above)</td>
<td>10 private APIs (same as above)</td>
</tr>
</tbody>
</table>

Patterns:
- district + random name
- district + random fruit name
- district + random street + random letters
- district + random street + specific letters
- **district + specific Chinese characters**
- district + specific last name + random first name
- district + random street + specific Chinese characters
- **district + random street + specific last name + random first name**
Hidden criminal infrastructures

Complicity of roles
Conclusion

Lesson learnt

• Chatbot is effective to study the cybercrime which are highly rely on crowdsourcing
• Account trading lies at the center of the fraud ecosystem, more effort should be put to mitigate the fraud account threats

Future work

• The current implementation of Aubrey is simple while effective;
• more complicated conversation (jargon identification), larger open domain corpora, hybrid model with human analyst involvement

https://sites.google.com/view/aubreychatbot
Thank you!
Discussion

• **Scope**
  • collected threat intel. is related to Chinese e-commerce platforms

• **Generalization**
  • with target intel. and domain-specific corpora, Aubrey can be re-trained to chat with other roles (drug dealers etc.) and languages

• **Impact**
  • fraud-related artifacts can be used as ground truth
  • fix exposed private APIs to raise the bar for automated abuse
  • stop fraudulent activities at the early stage
FSM for fake account trading
FSM for SIM farm and fake order operation

FSM for SIM farm

FSM for fake order operation
Knowledge source extension

Questions for miscreants

Answers to miscreants

Q&A

IM group chats + Forum discussions

similar as seed questions

extract Q&A pairs

candidate Q&A pairs
Data collection

• Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of raw data</th>
<th># of dialog pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seed conversation</td>
<td>800</td>
<td>200</td>
</tr>
<tr>
<td>IM group discussion</td>
<td>1 Million</td>
<td>50,000</td>
</tr>
<tr>
<td>Forum discussion</td>
<td>135,000</td>
<td>700,000</td>
</tr>
</tbody>
</table>
Evaluation

• Role identification classifier
  • Ground truth: 500 upstream, 180 downstream, 3,000 unrelated actors
  • Unknown set: 20,265 IM group members (from 150 IM groups)
  • Effectiveness:
    upstream: 87.0% precision, 91.2% recall
    downstream: 81.1% precision, 95.6% recall
    upstream actor: 89.0% precision, 92.8% recall
    overall: 86.2% F1 score

1,044 SIM farmers, 700 account merchants, 2,648 fraud order ops

• Accuracy
  • 545 chat attempts, 470 responded (185 SIM farmers, 130 account merchants, 155 fraud order operators);
  • one questioned Aubrey
  • 97.4% (458) accuracy
Effectiveness

CDF of interaction round per miscreant

CDF of interaction round for intel. gathering

52%
Case study

Revenue = sales * price = $48K/month

Account inventory and price tracking