Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale

Michele Campobasso & Luca Allodi

Security Group, Mathematics and Computer Science
An innovative underground market for user impersonation at scale

Operating under a new threat model affecting victims worldwide: **Impersonation-as-a-Service**[1]

Open gaps:
- Threat size
- Market revenue
- Attacker preferences → *which type of users are most at risk?*


Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
An innovative underground market for user impersonation at scale
Operating under a new threat model affecting victims worldwide: Impersonation-as-a-Service

Open gaps:
- Threat size
- Market revenue
- Attacker preferences → which type of users are most at risk?

A unique opportunity to directly measure supply & demand to estimate attacker preferences & market size (no proxy – e.g. user feedback)


Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
Challenges and solutions to accurately measure Genesis Market’s activity

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted access via invite</td>
<td>Infiltration in affiliated community to obtain 6 invites for 6 accounts</td>
</tr>
<tr>
<td>Anti-crawler mechanisms</td>
<td>Browser instrumentation + throttling + 24h sampling + splitting task among 6 crawlers</td>
</tr>
<tr>
<td>Data censored due to connectivity limitations (market availability &amp; TOR network congestion)</td>
<td>Retry crawling + collect summary data + statistical evaluation to estimate &amp; recreate missing data</td>
</tr>
<tr>
<td>High dimensionality &amp; noisy data</td>
<td>Aggregation of effects with dimensionality reduction, sales prediction model accounting for attacker decisions based on the daily supply</td>
</tr>
</tbody>
</table>
Challenges and solutions to accurately measure Genesis Market’s activity

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted access via invite</td>
<td>Infiltration in affiliated community to obtain 6 invites for 6 accounts</td>
</tr>
<tr>
<td>Anti-crawler mechanisms → tradeoff between data completeness &amp; stealth (risking ban)</td>
<td>Browser instrumentation + throttling + 24h sampling + splitting task among 6 crawlers</td>
</tr>
</tbody>
</table>
Challenges and solutions to accurately measure Genesis Market’s activity

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted access via invite</td>
<td>Infiltration in affiliated community to obtain 6 invites for 6 accounts</td>
</tr>
<tr>
<td>Anti-crawler mechanisms $\rightarrow$ tradeoff between data completeness &amp; stealth (risking ban)</td>
<td>Browser instrumentation + throttling + 24h sampling + splitting task among 6 crawlers</td>
</tr>
<tr>
<td>Data censored due to connectivity limitations (market availability &amp; TOR network congestion)</td>
<td>Retry crawling + collect summary data + statistical evaluation to estimate &amp; recreate missing data</td>
</tr>
</tbody>
</table>
Challenges and solutions to accurately measure Genesis Market’s activity

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted access via invite</td>
<td>Infiltration in affiliated community to obtain 6 invites for 6 accounts</td>
</tr>
<tr>
<td>Anti-crawler mechanisms → tradeoff between data completeness &amp; stealth (risking ban)</td>
<td>Browser instrumentation + throttling + 24h sampling + splitting task among 6 crawlers</td>
</tr>
<tr>
<td>Data censored due to connectivity limitations (market availability &amp; TOR network congestion)</td>
<td>Retry crawling + collect summary data + statistical evaluation to estimate &amp; recreate missing data</td>
</tr>
</tbody>
</table>

Inc. waking up at night to check if the crawler is working 😤
Challenges and solutions to accurately measure Genesis Market’s activity

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted access via invite</td>
<td>Infiltration in affiliated community to obtain 6 invites for 6 accounts</td>
</tr>
<tr>
<td>Anti-crawler mechanisms → tradeoff between data completeness &amp; stealth (risking ban)</td>
<td>Browser instrumentation + throttling + 24h sampling + splitting task among 6 crawlers</td>
</tr>
<tr>
<td>Data censored due to connectivity limitations (market availability &amp; TOR network congestion)</td>
<td>Retry crawling + collect summary data + statistical evaluation to estimate &amp; recreate missing data</td>
</tr>
<tr>
<td>High dimensionality &amp; noisy data</td>
<td>Aggregation of effects with dimensionality reduction, sales prediction model accounting for attacker decisions based on the daily supply</td>
</tr>
</tbody>
</table>

Inc. waking up at night to check if the crawler is working 😤
Measuring attacker preferences from an active market

1. DATA COLLECTION AND ENRICHMENT

- Profiles appearance - today
  - 30-90 min

- Profiles survived - previous days
  - 15-45 min

2. ORTHOGONALIZATION AND TIME WINDOW SELECTION

- GLMM sales prediction model (random effect daily supply) from 101/107 days with complete data
- Predict sales for the 6 days with no information on sales
- MonteCarlo sim. to recreate the 161 - 107 = 54 missing days
- Dimensionality reduction (MFA) on appeared profiles
  - Dimensions are linear combinations of variables (“profile class”)
- First day of sales only – 101/107 days with complete data
  - 57% total sales (only 6 days for which we cannot measure sales)

3. SALES PREDICTION AND LISTING RECONSTRUCTION

- 3 crawlers sampling 25% of appeared profiles in the last 24h at MSK midnight
- 3 crawlers checking “survived” profiles over the next 6 days
- Complete info for 107/161 days
  - 12'149 profiles with detailed info
- Data enriched with GDP/capita based on country of origin
- Classification of available credentials w.r.t. website purpose (moneytransfer, social, commerce, …)

4. MARKET DATA ANALYSIS

Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
Measuring attacker preferences from an active market

- Data collection from Jan 21\textsuperscript{st} to Jun 30\textsuperscript{th} 2021 (161 days)
Measuring attacker preferences from an active market

- Data collection from Jan 21\textsuperscript{st} to Jun 30\textsuperscript{th} 2021 (161 days)
- 3 crawlers sampling 25\% of appeared profiles in the last 24h at MSK midnight
- 3 crawlers checking ”survived” profiles over the next 6 days
Measuring attacker preferences from an active market

- Data collection from Jan 21\textsuperscript{st} to Jun 30\textsuperscript{th} 2021 (161 days)
- 3 crawlers sampling 25% of appeared profiles in the last 24h at MSK midnight
- 3 crawlers checking "survived" profiles over the next 6 days
- Complete info for 107/161 days $\rightarrow$ 12’149 profiles with detailed info
- Data enriched with GDP/capita based on country of origin
- Classification of available credentials wrt website purpose (moneytransfer, social, commerce, ...)

Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
Measuring attacker preferences from an active market

- Dimensionality reduction (MFA) on appeared profiles
- Dimensions are linear combinations of variables ("profile class")

1. DATA COLLECTION AND ENRICHMENT
   - Profiles appearance - today
   - Profiles survived - previous days
   - Complete info for 107/161 days

2. ORTHOGONALIZATION AND TIME WINDOW SELECTION
   - MFA

Data collection from Jan 21st to Jun 30th 2021 (161 days)
- 3 crawlers sampling 25% of appeared profiles in the last 24h at MSK midnight
- 3 crawlers checking "survived" profiles over the next 6 days
- Complete info for 107/161 days
- Data enriched with GDP/capita based on country of origin
- Classification of available credentials w.r.t. website purpose (moneytransfer, social, commerce, …)
Measuring attacker preferences from an active market

- Dimensionality reduction (MFA) on appeared profiles
  - Dimensions are linear combinations of variables (“profile class”)

- First day of sales only → 101/107 days with complete data – 57% total sales (only 6 days for which we cannot measure sales)
Measuring attacker preferences from an active market

- GLMM sales prediction model (random effect → daily supply) from 101/107 days with complete data
- Predict sales for the 6 days with no information on sales
Measuring attacker preferences from an active market

- GLMM sales prediction model (random effect → daily supply) from 101/107 days with complete data
- Predict sales for the 6 days with no information on sales
- Montecarlo sim. to recreate the 161-107=54 missing days

Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
Measuring attacker preferences from an active market

Finally (!), we can study (spoiler alert):

• Attackers’ purchasing decisions

• Scale of the threat

• Market revenues

• The relationship between findings and Woods & Böhme’s risk model
Rich nations are targeted, nuanced profiles selection

Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
Rich nations are targeted, nuanced profiles selection

- +60% supply from EU, 12% NA
- Supply ~matches demand in NA & OCE (NA 4x more preferred than EU)
Rich nations are targeted, nuanced profiles selection

- +60% supply from EU, 12% NA
- Supply ~matches demand in NA & OCE (NA 4x more preferred than EU)

Dim8: Profiles from wealthier countries (positive sign wdi) rich in stolen cookies from Chrome (positive sign n_cookies_chrome & n_chrome)
Rich nations are targeted, nuanced profiles selection

- +60% supply from EU, 12% NA
- Supply ~matches demand in NA & OCE (NA 4x more preferred than EU)

Know Your Cybercriminal: Evaluating Attacker Preferences by Measuring Profile Sales on an Active, Leading Criminal Market for User Impersonation at Scale
Rich nations are targeted, nuanced profiles selection

- +60% supply from EU, 12% NA
- Supply ~matches demand in NA & OCE (NA 4x more preferred than EU)

Dim8: Profiles from wealthier countries (positive sign wdi) rich in stolen cookies from Chrome (positive sign n_cookies_chrome & n_chrome)

Dim2: Profiles from poorer countries (negative sign wdi) exhibiting older OSs (positive sign win7/8)

Dim8: are more likely to be sold (positive coefficient)
Rich nations are targeted, nuanced profiles selection

- +60% supply from EU, 12% NA
- Supply ~matches demand in NA & OCE (NA 4x more preferred than EU)

Dim8: Profiles from wealthier countries (positive sign wdi) rich in stolen cookies from Chrome (positive sign n_cookies_chrome & n_chrome)
Dim2: Profiles from poorer countries (negative sign wdi) exhibiting older OSs (positive sign win7/8)

Dim8: are more likely to be sold (positive coefficient)
Dim2: are less likely to be sold (negative coefficient)
A lucrative, worldwide scale threat

Offered profiles median prices and volume

First world regions are main targets, Europe first
Ex-URSS countries are not included
A lucrative, worldwide scale threat

Offered profiles median prices and volume

First world regions are main targets, Europe first
Ex-URSS countries are not included

Supply and demand (i.e., actual affected users) and revenues

Scaled up numbers (accounting for sampling & data censorship):
A lucrative, worldwide scale threat

Offered profiles median prices and volume

Supply and demand (i.e., actual affected users) and revenues

First world regions are main targets, Europe first
Ex-URSS countries are not included

Scaled up numbers (accounting for sampling & data censorship):

- Up to 3’800 new listed profiles daily (600 on average)
A lucrative, worldwide scale threat

Offered profiles median prices and volume

Supply and demand (i.e., actual affected users) and revenues

First world regions are main targets, Europe first
Ex-URSS countries are not included

Scaled up numbers (accounting for sampling & data censorship):
- Up to 3’800 new listed profiles daily (600 on average)
- Up to 430 profiles sold (actual people attacked) daily (avg 125)
A lucrative, worldwide scale threat

First world regions are main targets, Europe first
Ex-URSS countries are not included

Offered profiles median prices and volume

Supply and demand (i.e., actual affected users) and revenues

Scaled up numbers (accounting for sampling & data censorship):
- Up to 3’800 new listed profiles daily (600 on average)
- Up to 430 profiles sold (actual people attacked) daily (avg 125)
- Estimated daily revenues: avg 3’000 – 4’000$, max: 7’200 – 16’400$
A lucrative, worldwide scale threat

Offered profiles median prices and volume

Supply and demand (i.e., actual affected users) and revenues

First world regions are main targets, Europe first
Ex-URSS countries are not included

Scaled up numbers (accounting for sampling & data censorship):
- Up to 3’800 new listed profiles daily (600 on average)
- Up to 430 profiles sold (actual people attacked) daily (avg 125)
- Estimated daily revenues: avg 3’000 – 4’000$, max: 7’200 – 16’400$
- Estimated yearly market revenues: 1.2M – 1.6M$
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Risk model
[SoK - Quantifying Cyber Risk; Daniel W. Woods, Rainer Böhme]
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):

- expected wealth (country WDI)
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):

- expected wealth (country WDI)
- Technical details (mainly browsers)
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):

- expected wealth (country WDI)
- Technical details (mainly browsers)

Stakeholders could monitor market activity to:
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):

- **expected wealth (country WDI)**
- **Technical details (mainly browsers)**

Stakeholders could monitor market activity to:

- **evaluate exposure (market shows affected websites, ...)**
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):

- expected wealth (country WDI)
- Technical details (mainly browsers)

Stakeholders could monitor market activity to:

- evaluate exposure (market shows affected websites, …)
- prepare to attacks (revoke creds, mandate 2FA for a week, inform relevant SOC, …)
Attacker preferences to inform Woods & Böhme’s cyber risk model

Threat is a consequence of the attackers’ decisions.

Purchase decisions are mostly influenced by (but there’s more):

- expected wealth (country WDI)
- Technical details (mainly browsers)

Stakeholders could monitor market activity to:

- evaluate exposure (market shows affected websites, ...)
- prepare to attacks (revoke creds, mandate 2FA for a week, inform relevant SOC, ...)

We should always consider how attackers think to better evaluate risks for the final user.
A tale of a market takedown

Market takedown in April.

2 months later, Genesis Market has been sold, (inc. infrastructure & impersonation software).

Likely to see another similar market in the near future.

Check if you were/are a victim: [https://politie.nl/checkyourhack](https://politie.nl/checkyourhack)