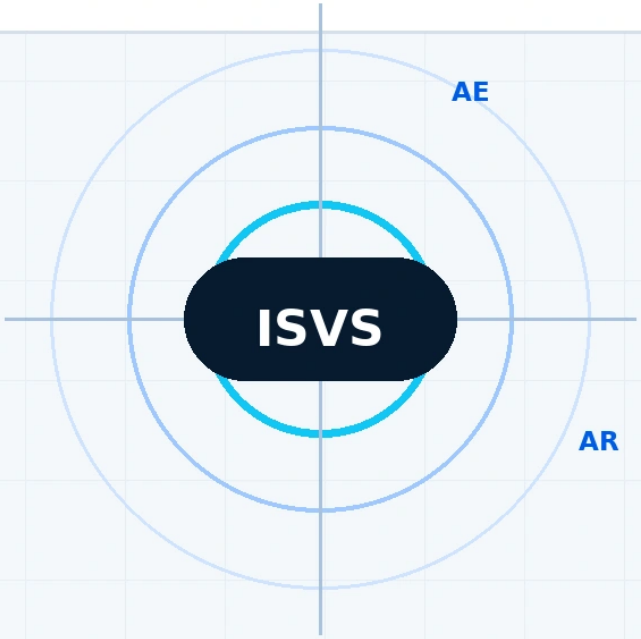
Threat Hunter
BUSINESS SECURITY RESEARCH

THREAT HUNTER RESEARCH

BUSINESS SECURITY BLUE TEAM ASSESSMENT STANDARD WHITE PAPER

2024 EDITION

A quantitative framework for evaluating authorized business-security attack simulations



Published by Threat Hunter

Full English Translation | Original publication: 12 March 2024

About this English edition

PUBLICATION

Business Security Blue Team Assessment Standard White Paper (2024 Edition)

PUBLISHER

Threat Hunter

ORIGINAL PUBLICATION DATE

12 March 2024

TRANSLATION NOTE

This is a complete English translation of the substantive content in the 31-page Chinese publication. The original formulae, assessment values, scenario results, tables, figure meanings and publication date are preserved. Figure labels have been redrawn in English for legibility, and the Threat Hunter brand name is used consistently throughout.

The cover, repeated table of contents and purely decorative source pages are not reproduced as body text. Minor terminology adaptations improve clarity for international readers without changing the underlying methodology. In case of ambiguity, consult the original Chinese publication.

SOURCE RECORD

Original Chinese PDF: 31 pages. Source SHA-256: 31dd305cc17686ea92261f03ada73aa66ec76740b3e1794b99273e7c38d3d1f5. Translation source and generated assets are maintained in the Threat Hunter website repository.

Contents

This page is the complete English translation of the 31-page Chinese report. The original publication date, substantive sections, case studies, scoring tables and appendix are preserved. "Blue Team" in this report means an authorized, controlled business-security adversarial assessment.

Foreword

The cybercriminal and illicit grey-market ecosystem has evolved into an organized, automated, and highly interconnected industrial network, with tightly coordinated upstream and downstream participants.

According to industry estimates, cybercriminal and grey-market activity causes nearly RMB 100 billion in losses to internet companies and offline industries every year. Participants in this ecosystem disguise themselves as legitimate business requests while continuously eroding enterprise value. Examples include promotion-abuse groups that capture new-user incentives, traffic-acquisition operations that exploit enterprise platforms to conduct malicious marketing and fraud at scale, and traffic-manipulation operations that undermine recommendation and billing rules. As attack models mature and operating models become more reproducible, business security problems are becoming increasingly prominent.

Enterprises commonly face the following challenges when addressing business security:

- **Information asymmetry between attackers and defenders.** Enterprises often lack a deep understanding of criminal techniques and their evolution, resulting in delayed detection, delayed response, and long remediation cycles.
- **Difficulty evaluating control effectiveness.** After a control is deployed, it can be difficult to assess its overall effectiveness or promptly identify new evasion techniques.
- **Absence of an assessment framework.** Unlike foundational cybersecurity, business security has no clear-cut boundary. The severity of many foundational security issues can be evaluated using standards such as the OWASP Top 10. Business security, however, involves complex scenarios and must also account for user experience and user activity. The industry has therefore lacked an effective, practical assessment framework.

Unlike foundational security, where the goal may be to block an attack completely, the objective of business security is usually not to eliminate every attack. Instead, the objective is to contain attacks within an acceptable range so that business operations can continue normally and the business can grow in a healthy way.

Consider promotion abuse. An online promotion intended to clear physical inventory, or a restaurant discount coupon that requires in-store redemption, may tolerate a relatively high share of incentive-abuse participants. The business security objective might be to keep them below 50%. In contrast, an acquisition campaign aimed at a specific customer segment is intended to direct subsidies to genuine users and therefore has a much lower tolerance for abuse. Genuine users may convert into long-term customers, whereas the conversion rate for incentive-abuse accounts is nearly zero. If abuse is too prevalent, the enterprise incurs substantial opportunity costs; the governance objective might therefore be to keep the share below 10%. Even within promotion campaigns, different campaign types and rules lead to different assessment objectives.

Enterprises therefore need a business security assessment framework that can describe, in a quantitative and systematic way, both the harm caused by attacks and the effects of the controls implemented in response.

To address this need, Threat Hunter published China's first *Business Security Blue Team Assessment Standard White Paper* in 2020, filling a long-standing gap in the industry's ability to assess attack impact and control effectiveness. Four years later, Threat Hunter revised and updated the original standard based on challenges encountered during implementation and on several years of practical Blue Team assessment experience with customers across multiple industries. The result is the *Business Security Blue Team Assessment Standard White Paper (2024 Edition)*.

1. Business Security Blue Team Assessment Standard

1.1 Interaction Security Vulnerability Score (ISVS)

The Interaction Security Vulnerability Score (ISVS) evaluates business risk from an attacker's perspective. A specific enterprise business object is designated as the assessment target. The assessment team uses attack methods observed in the cybercriminal and grey-market ecosystem to conduct controlled simulated attacks against that target, reconstructs the attack process, and evaluates the harm that the attack could cause based on the final test results.

The ISVS formula is:

```
ISVS = Max(Attack Efficiency (AE) x Achievement Rate (AR))
```

Attack Efficiency (AE)

The method used to determine Attack Efficiency varies with the assessment target and intended objective. It is generally evaluated from two perspectives: material-resource acquisition efficiency and technical-adversarial efficiency.

Material resources include phone numbers, accounts, IP addresses, devices, identity-verification resources, and other inputs required during an attack. The lower the cost of obtaining these materials, the higher the Attack Efficiency; the higher the cost, the lower the Attack Efficiency.

Technical factors include the techniques used to break application protections or evade risk-control restrictions, such as software reverse engineering, device-identity modification, location modification, and face fabrication. The lower the threshold for attackers to apply these techniques successfully, the higher the Attack Efficiency; the higher the threshold, the lower the Attack Efficiency.

> *See the appendix for reference levels used when assigning Attack Efficiency.*

Achievement Rate (AR)

Achievement Rate measures how much of the intended attack objective a given attack plan achieves in the final test. If the result meets or exceeds the intended objective, AR is 1.

For a quantitative objective, suppose the goal is 10,000 successful attacks within a specified period and the test achieves 8,000. The Achievement Rate is 8,000 / 10,000 = 0.8.

For a non-quantitative objective, suppose the goal is to bypass a target's face-recognition control. If the test bypasses the control successfully, AR is 1; otherwise, AR is 0.

If an objective can be achieved only under constrained conditions - for example, only on an older Android version - AR may be assigned a value between 0 and 1 according to the actual circumstances.

AE and AR are each scored in the interval [0, 1]. Multiplying them produces an ISVS value that is also in the interval [0, 1].

An assessment may use multiple attack plans. The overall ISVS takes the maximum score among them, following the weakest-link principle. A higher ISVS means that the assessed business object is more vulnerable to the corresponding attack plan and that the potential harm would be greater if criminal actors used that plan against the business.

FIGURE 1 | ISVS PLAN DISTRIBUTION

The highest plan score determines overall ISVS

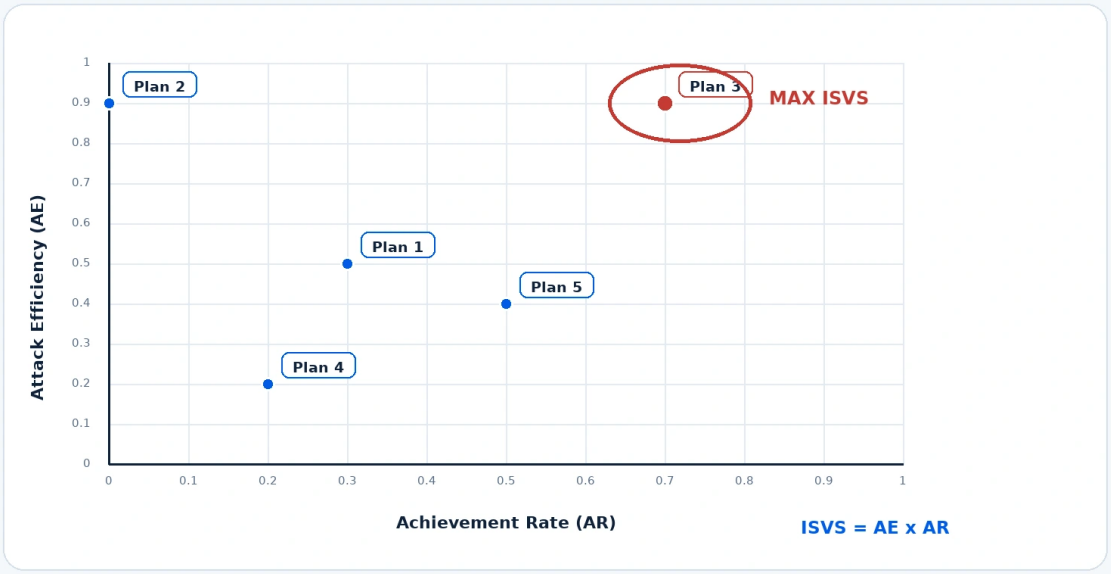


Figure 1. Example ISVS distribution across five attack plans

Source: Threat Hunter, original Chinese report, page 8

1.2 How to interpret an ISVS result

The purpose of ISVS is not simply to pursue the lowest possible score. It is to reflect the current state of business security objectively from an attacker's perspective, expose weaknesses promptly, and drive remediation and governance.

1.2.1 Baseline comparison

As noted above, business security is not intended to eliminate every attack; it is intended to contain attacks within an acceptable range. If an enterprise defines 1,000 attempts using a widely accessible attack method (AE = 1), fewer than 200 successes as controlled, and more than 500 successes as out of control, it can establish two ISVS reference values:

- **Controlled baseline:** 0.2
- **Out-of-control baseline:** 0.5

A score in [0, 0.2) indicates that the attack has been contained within the controlled range. A score in (0.5, 1] indicates that the situation is out of control and that security capabilities need to be strengthened urgently.

FIGURE 2 | BASELINE COMPARISON

Controlled and out-of-control baseline regions

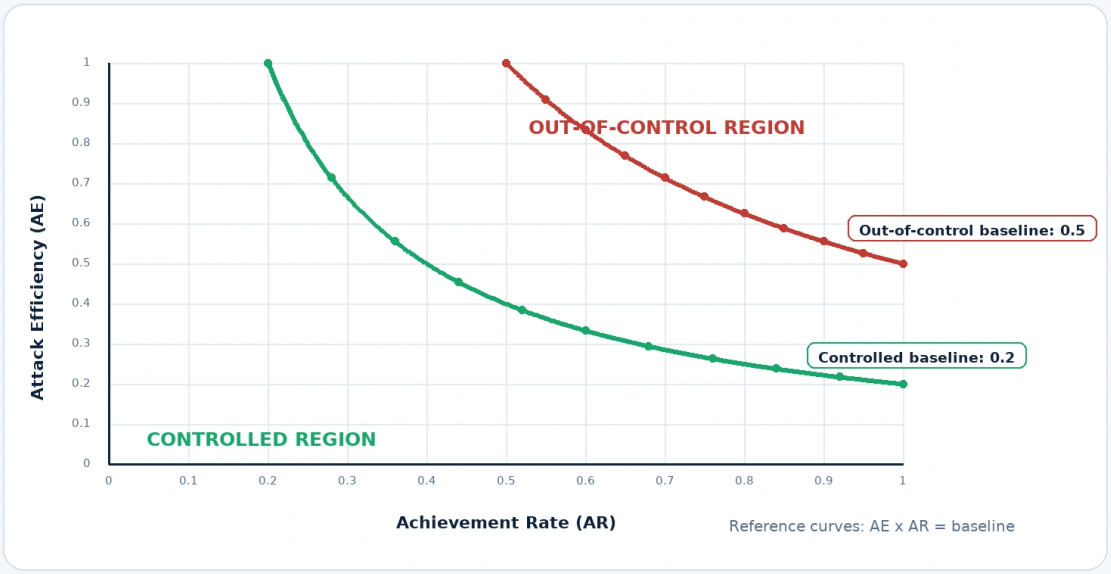


Figure 2. Controlled and out-of-control ISVS regions

Source: Threat Hunter, original Chinese report, page 9

1.2.2 Longitudinal comparison

When the assessment object and intended objective remain unchanged, the ISVS can still change as the enterprise's business security posture evolves and criminal attack methods are upgraded. Business Security Blue Team assessments should therefore be conducted periodically. Comparing ISVS values over time shows whether security investment is producing real results and whether new weaknesses have emerged.

FIGURE 3 | LONGITUDINAL COMPARISON

Periodic assessment exposes change over time

Figure 3. Longitudinal ISVS comparison across assessment rounds

Source: Threat Hunter, original Chinese report, page 10

1.2.3 Horizontal comparison

Horizontal assessment compares an object with similar businesses in the same industry and uses ISVS to reflect the object's relative security posture.

If the object's ISVS is higher than that of comparable businesses, security capabilities should be strengthened as soon as possible because a relatively weak business will naturally become a priority target for criminal actors.

FIGURE 4 | INDUSTRY-WIDE COMPARISON

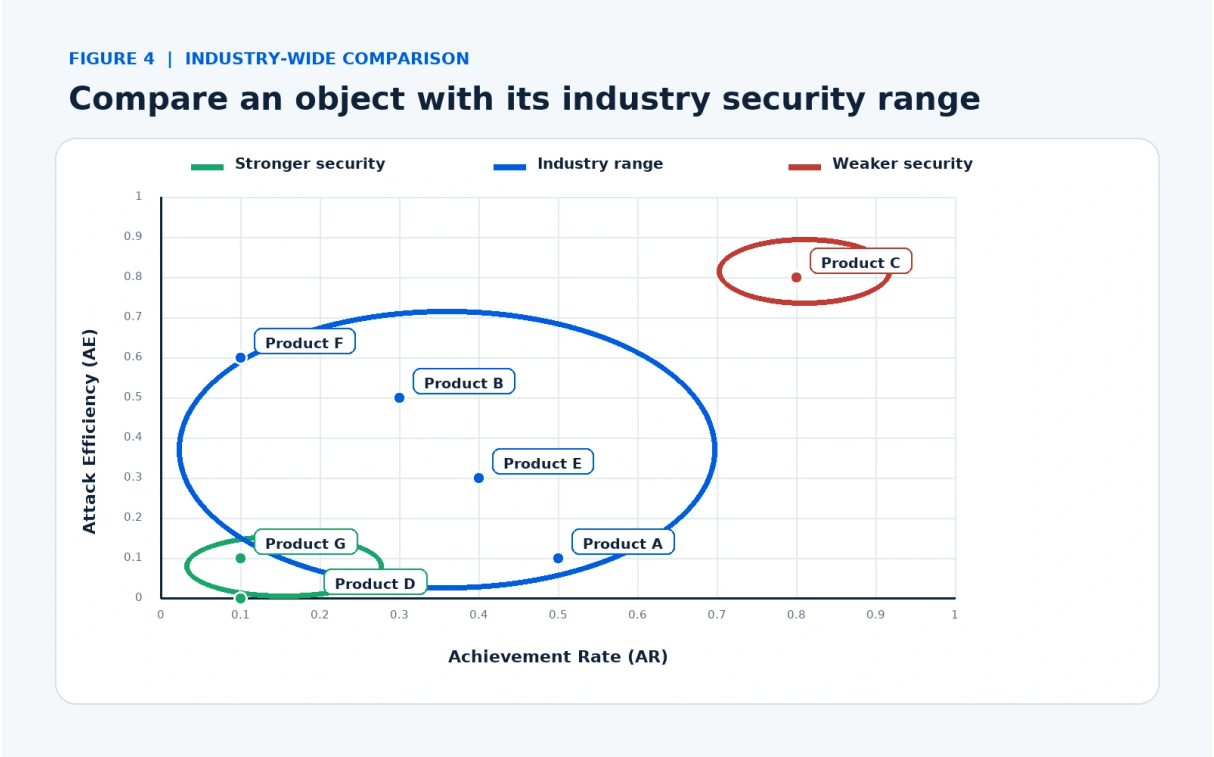


Figure 4. Horizontal ISVS comparison within an industry

Source: Threat Hunter, original Chinese report, page 11

2. Business Security Blue Team Assessment Cases

2.1 Install-fraud assessment case

Install fraud commonly occurs in marketing and user-acquisition programs. Some promotion channels collude with criminal actors to fabricate app-install volumes and defraud enterprises of promotion fees.

2.1.1 Define the assessment object and attack objective

- **Assessment object:** a social application
- **Intended objective:** conduct testing for one week, with each attack plan producing an average of 2,000 successful fraudulent installs per day

2.1.2 Develop attack plans

Based on intelligence gathered through criminal-ecosystem research, install-fraud techniques generally follow two attack paths from the attacker's perspective:

1. Use scripts on physical phones or emulators to simulate application installation and launch. A device-control system operates many devices in batches, while device-identity modification changes IMEI, Android ID, MAC address, and other parameters to make each installation appear to come from a new device.
2. Reverse engineer the application's interface protocol and write an automated program that fabricates installation and launch requests. False device information is inserted into request parameters so that the business backend is deceived without the application actually being installed.

Four attack plans were developed from these two paths:

1. Use emulators to simulate multiple mobile devices.
2. Use physical phones and software-based device-identity modification to simulate new devices.
3. Use physical phones and custom-ROM device-identity modification to simulate new devices.
4. Reverse engineer the application interface protocol and directly fabricate installation and launch requests.

2.1.3 Execute the attack plans and evaluate ISVS

Attack plan	Assessment result	AE	AR	ISVS
Plan 1	The application had strong emulator-detection capabilities. Multiple emulators were tested, but none could launch the application normally. Attempts to deliberately remove emulator characteristics using methods known in criminal communities also failed.	0	N/A	**0**
Plan 2	The application had strong client-side defenses. Multiple software-based device-modification tools were tested, but the application detected injection, Hook, or similar traces and refused to launch normally.	0	N/A	**0**
Plan 3	A custom ROM modified low-level system source code to change device information, preventing the application from defending against it effectively. The application could be installed and launched, and each device modification was recognized as an installation on a new device. During the assessment period, however, custom-ROM modification was not yet widely used by criminal actors and required dedicated hardware. The average number of fraudulent installs completed per day was 1,200.	0.5	1,200 / 2,000 = 0.6	**0.3**
Plan 4	Protocol-based attacks were common, but the application encrypted its interface protocol, so packet capture exposed only encrypted content and could not be used directly to fabricate requests. The application also protected its code, preventing direct decompilation of the encryption algorithm or direct invocation of the encryption function. A senior reverse-engineering engineer required approximately one month to break the protocol algorithm, representing low technical-adversarial efficiency for criminal actors. After the protocol was broken, more than 2,000 fraudulent installs could be completed per day.	0.1	1	**0.1**

Across Plans 1 through 4, the assessed object's overall ISVS was **0.3**.

The assessment conclusion was:

> *The assessed object has strong defenses against install fraud. Criminal actors cannot use emulators or software-based device modification to simulate new devices and create fraudulent installs. Fabricating protocol requests also presents a high technical threshold, and periodic algorithm changes can further constrain this form of attack. The application does not, however, detect custom-ROM device modification. Because criminal actors are expected to use this approach more widely, targeted defenses should be introduced as soon as possible.*

2.2 Face-recognition bypass assessment case

Face-recognition bypass is most relevant to critical identity-verification scenarios. For example, after criminals obtain a victim's account credentials, they may still need to bypass face recognition before they can log in from a different location, transfer funds, or perform other sensitive actions.

2.2.1 Define the assessment object and attack objective

- **Assessment object:** a financial application
- **Intended objective:** successfully bypass the application's face-recognition control and log in to an account as someone other than the legitimate user

2.2.2 Develop attack plans

Based on intelligence gathered through criminal-ecosystem research, bypassing face recognition involves two key steps: generating a fabricated face video and controlling the mobile-phone camera so that it plays that video.

Two methods were used to generate a face video:

1. **Face animation:** use software such as CrazyTalk to generate a video from a victim's photograph, including blinking, nodding, and head-shaking movements.
2. **Face replacement:** use software such as DeepFaceLab to replace the face in a prerecorded video with the victim's face.

Three methods were used to present the face video to the mobile application:

1. **Screen replay:** play the face video on a computer and point the phone camera at the display.
2. **Hijack the mobile camera:** use a specially modified phone designed to bypass face checks and hijack the camera video stream.
3. **Use a cloud-phone virtual camera:** use a cloud phone's virtual-camera function to play a designated video.

Combining the two video-generation methods with the three presentation methods produced six attack plans:

1. Face animation + screen replay
2. Face animation + mobile-camera hijacking
3. Face animation + cloud-phone virtual camera
4. Face replacement + screen replay
5. Face replacement + mobile-camera hijacking
6. Face replacement + cloud-phone virtual camera

2.2.3 Execute the attack plans and evaluate ISVS

Attack plan	Assessment result	AE	AR	ISVS
Plan 1	A video produced through face animation could not pass detection.	N/A	0	**0**
Plan 2	Same result as Plan 1.	N/A	0	**0**
Plan 3	Same result as Plan 1.	N/A	0	**0**
Plan 4	A screen-replay attack using the phone camera could not pass detection.	N/A	0	**0**
Plan 5	Hijacking the camera and playing a face-replacement video passed detection and allowed a successful login. The method required purchasing a dedicated modified phone and was not guaranteed to be stable.	0.5	1	**0.5**
Plan 6	Playing a face-replacement video through a cloud phone's virtual camera passed detection and allowed a successful login. The barrier to using a cloud phone was low, but this method had not yet been adopted by criminal actors during the assessment period.	0.2	1	**0.2**

Across Plans 1 through 6, the assessed object's overall ISVS was **0.5**.

The assessment conclusion was:

> *The assessed object has a security blind spot in its defenses against face-recognition bypass. Videos generated using face-animation technology could not bypass detection. Videos generated using face-replacement technology, however, could bypass detection when combined with either camera hijacking on a specially modified phone or a cloud-phone virtual camera. Malicious exploitation by criminal actors could cause significant asset losses for users, so targeted defenses should be introduced as soon as possible.*

3. Types of Business Security Blue Team Assessment

The cybercriminal and grey-market ecosystem has become organized, automated, and industrialized, with tightly coordinated upstream and downstream participants. As attack models mature and operating models become easier to reproduce, business security problems become more prominent. Enterprises face challenges such as information asymmetry and difficulty measuring control effectiveness. Business Security Blue Team assessment services emerged in response.

Business Security Blue Team testers simulate real criminal actors. Within a controlled scope, they use resources, tools, and methods observed in the real criminal ecosystem to conduct non-destructive attack simulations and reconstruct attack details. This helps an enterprise understand:

- **The state of the criminal value chain.** This includes the input costs, potential returns, overall attack scale within the value chain, and the scale of criminal activity targeting the enterprise itself. These insights help an enterprise optimize its response to attacks from multiple perspectives, including cost; improve defensive efficiency; and raise the attacker's barrier to entry.
- **Details of attack methods.** First-hand data and attack-method details help an enterprise adjust its risk-control policies and rules in a timely and appropriate manner.

Business Security Blue Team assessments currently fall into the following main types.

3.1 Foundational assessment

A foundational assessment is suitable for the early and middle stages of business security development. At this stage, enterprises often have limited understanding of the cybercriminal and grey-market ecosystem. They may not know whether their business is being attacked, whether attacks have succeeded, or what harm those attacks could cause. A professional assessment is therefore needed to build an informed view of the threat and guide targeted security improvements.

In addition to the Business Security Blue Team assessment report, Threat Hunter can provide a related criminal-value-chain research report according to the customer's needs, helping the customer understand the ecosystem in greater depth.

3.2 Periodic assessment

A periodic assessment is suitable for long-running business functions or promotion campaigns. Once business security has reached the middle stage of development, continued changes in both the business and the criminal ecosystem make recurring assessments necessary. They show whether the current security posture has improved or deteriorated, evaluate the specific effects of security controls, and reveal new blind spots promptly.

3.3 Industry-wide comparative assessment

An industry-wide comparative assessment specifies identical or comparable test items for the same scenario, selects different companies in the same industry, performs the same assessments, and compares the results horizontally.

This type of assessment is subject to limitations. Any test item that requires authorization cannot be performed against an organization for which authorization has not been obtained.

4. Business Security Blue Team Assessment Scenarios

4.1 Promotion-abuse scenarios

Subsidies, new-user red envelopes, referral incentives, and cash-withdrawal campaigns are common enterprise marketing techniques and common targets for criminal actors seeking economic benefit.

Different campaigns and objectives have different exposure to promotion abuse and different levels of risk tolerance, so their intended assessment objectives also differ. Typical assessment items include the following.

1. Bulk acquisition of designated coupons

- **Test item:** obtain and monetize RMB 5, RMB 10, or RMB 15 new-user coupons from a platform.
- **Intended objective:** complete 100 account registrations within one hour, with each account able to obtain and use a coupon successfully.

2. Group-purchase campaigns

- **Test item:** assess an existing-user-refers-new-user group-purchase campaign on a platform.
- **Intended objective:** complete 30 group purchases within one hour. An attempt is considered successful when the order enters fulfillment, taking into account the availability of channels for monetizing physical goods.

3. Referral-based acquisition and cash withdrawal

- **Test item:** assess a platform campaign in which users invite friends to earn cash.
- **Intended objective:** use an existing account to recruit new users, obtain the maximum reward, execute the process in batches, and withdraw the reward successfully.

Additional considerations include:

1. A marketing campaign may run in an application, a mini program, or an H5 web page. Attack Efficiency varies across endpoints.
2. Criminal actors ultimately need a way to monetize promotion abuse. According to the customer's requirements, the number of successfully monetized rewards can therefore be used as the attack objective.

4.2 Business-metric manipulation scenarios

Business-metric manipulation commonly affects social, content, and e-commerce platforms. Criminal actors inflate views, clicks, plays, likes, comments, and other metrics associated with a designated object. This can be used either to obtain platform incentives or to improve ranking, exposure, and resource allocation, damaging the platform ecosystem.

Typical assessment items include the following.

1. Inflating views, clicks, or plays

- **Test item:** a designated video on a video platform.
- **Intended objective:** complete 10,000 video plays within one day without the backend risk-control system identifying them as false traffic.

2. Inflating likes, replies, or comments

- **Test item:** a designated post on a social platform.
- **Intended objective:** complete 500 replies within one day without the backend risk-control system identifying them as false traffic.

4.3 Advertising-traffic manipulation scenarios

Advertising-traffic manipulation has long been an illicit practice within the advertising industry. Depending on the settlement model, criminal actors fabricate impressions, clicks, downloads, installations, and other metrics to defraud advertisers of advertising expenditure. Advertisers do not always expect 100% of traffic to be genuine, so assessment items are mainly designed around two needs.

1. Constrain large-scale manipulation

The assessment determines whether criminal actors can use device-control systems, fabricated protocols, or other methods to create large volumes of false advertising impressions or clicks.

2. Compare channel quality

Assessment data is used to calculate the proportion of false traffic across the advertiser's delivery channels, providing an important basis for channel-quality comparison.

4.4 Face-recognition bypass scenarios

Face recognition is one of the most important forms of identity verification. It is used to ensure that sensitive business actions such as login, transfers, and payments are performed by the legitimate user. Once face recognition is bypassed, a highly trusted defensive mechanism has been defeated, making it difficult to prevent subsequent harm effectively.

Business Security Blue Team assessment of face-recognition bypass is therefore essential for financial, payment, and other applications closely linked to user assets.

As AI technologies such as ChatGPT continue to develop, fabricated face video will become increasingly realistic. This will remain a long-term adversarial process, so assessments need to be conducted periodically.

5. Business Security Blue Team Assessment Process

1. **Discuss assessment requirements.** Both parties discuss and document the assessment object, intended attack objectives, and related requirements.
2. **Prepare recommended test items.** The Blue Team assessment group produces a document containing recommended test items and assessment criteria according to the customer's needs.
3. **Agree on the object and objective.** Both parties discuss and confirm the assessment object and intended attack objective.
4. **Execute the contract and authorization agreement.**
5. **Conduct the assessment.**
6. **Accept and sign off the results.**

Appendix: Reference Levels for Attack Efficiency

A. Material-resource acquisition efficiency

A.1 Phone numbers

Acquisition condition	Attack Efficiency
The SMS verification process has brute-force, response-disclosure, or comparable vulnerabilities, allowing an arbitrary phone number to be entered.	High
Phone numbers acquired through criminal SMS-verification-code platforms can be used for registration, login, and business operations.	High
Phone numbers acquired through private criminal channels can be used for registration, login, and business operations.	Medium
Only phone numbers actively used by genuine individuals can be used for registration, login, and business operations.	Low
Detection and enforcement are separated, and an account registered with a malicious phone number is restricted only when it performs a malicious business action.	Low

A.2 Accounts

Acquisition condition	Attack Efficiency
Newly registered accounts have no restrictions and can perform sensitive business actions such as posting or replying in batches.	High
Newly registered accounts cannot perform sensitive business actions until they accumulate a required level of activity.	Medium
Automated account-farming can be used to accumulate account activity.	High
Automated account-farming cannot be used; accounts must be farmed manually or genuine-user accounts must be purchased.	Low
No verification is required when an account logs in from a new device or region.	High
SMS or similar verification is required when an account logs in from a new device or region.	Medium
A business license can be fabricated using image-editing software and used to register an enterprise account successfully.	High

A.3 IP addresses

Acquisition condition	Attack Efficiency
There are no controls for IP request frequency, common login-region IPs, or similar risk signals.	High
Proxy IPs or dial-up rotating IPs can bypass the risk-control system.	Medium
Proxy IPs or dial-up rotating IPs can bypass the risk-control system, but some attacks are identified or blocked.	Medium
Proxy IPs or dial-up rotating IPs have difficulty bypassing the risk-control system.	Low

A.4 Devices

Acquisition condition	Attack Efficiency
Common emulators can simulate multiple devices.	High
Multi-instance or application-cloning tools can simulate multiple devices.	High
Software-based device-identity modification can simulate multiple devices.	High
Custom-ROM device-identity modification can simulate multiple devices.	Medium
Hardware-based device-identity modification can simulate multiple devices.	Low

B. Technical-adversarial efficiency

B.1 Protocol reverse engineering and request fabrication

Reverse-engineering or fabrication condition	Attack Efficiency
Requests captured using a packet-capture tool can be replayed directly.	High
Common decompilation tools can directly restore the code related to protocol requests.	High
Code related to protocol requests is protected, but the protection is easy to break.	High
Code related to protocol requests is strongly protected and difficult to break, but the relevant function can be invoked directly.	Medium
Code related to protocol requests is strongly protected and difficult to break, and the relevant function cannot be invoked directly.	Low

B.2 Client-side defense

Adversarial condition	Attack Efficiency
Common frameworks such as Xposed or LSPosed can inject into the application process and conduct the attack.	High
Criminally modified injection frameworks can inject into the application process and conduct the attack.	Medium
Injection into a system process or service can be used to conduct the attack.	Medium
None of the preceding adversarial methods can be used to conduct the attack successfully.	Low

B.3 Bot-detection and CAPTCHA defense

Adversarial condition	Attack Efficiency
No bot-detection CAPTCHA is used, or only a simple image CAPTCHA is used.	High
A more complex slider or point-selection CAPTCHA is used, requiring an illicit CAPTCHA-solving service.	Medium
A device-control system, a macro or scripting tool such as Anjian Jingling or AutoJs, or browser automation such as WebDriver or Selenium can simulate human operation and complete the attack successfully.	High
The same automation approaches can simulate human operation, but some attacks are identified or blocked.	Medium
The same automation approaches have significant difficulty completing the attack successfully.	Low