silo
BY AUTHENTIC8

WHITE PAPER

Empowering Trust and Safety Teams

Protecting a brand's online communities and services from harmful content and inappropriate conduct is essential for maintaining trust. That's the focus of a trust and safety team. When analysts need to research potentially damaging or dangerous situations, investigations may require anonymous browsing and access to untrusted sites. Yet secure, anonymous access is getting tougher, and risky resources are often blocked. How can trust and safety investigators get the job done more effectively?

Here we'll take a closer look at the kind of work trust and safety teams do, the potential risks and how to empower teams to be successful.

The rising role of trust and safety

Trust and safety teams are playing an increasingly critical role in helping companies mitigate risk and preserve brand reputation. Typically, teams focus on key areas, such as:

- **Ensuring integrity of services:** promoting customer/user confidence by establishing and enforcing policies for acceptable use and compliance
- **Protecting users of online services:** preventing fraud and account takeovers or avoiding phishing scams
- **Protecting online services platforms:** managing security threats and protecting from brand misuse around their services and how they operate

Many organizations rely on automated systems to moderate content and user behavior. But simply reacting to flags isn't enough — often deeper investigations are required. For example, automation may incorrectly flag issues with legitimate users, and if you block those people arbitrarily, you risk ruining the "trust" part of trust and safety. Or a single flag might miss a bigger picture of persistent or organized misuse, where illegal activity needs to be investigated and relayed to law enforcement.

Investigations need to be safe as well, and minimizing risk is key. Best practice open source intelligence (OSINT) tradecraft helps ensure secure access and anonymity for online research, so investigations can never be traced back to an analyst or their organization. VPNs and private browsing — even parallel network infrastructure — are often not enough to protect trust and safety teams. And complex protocols make a tough job tougher. To work efficiently, analysts need tools that simplify access while providing greater control over how their online presence is attributed to them and their organization, so they can protect their research and ensure a successful investigation.

As the fraud landscape intensifies, companies are feeling pressured to build a strong team of investigators. Analysts with backgrounds in military/intelligence, cybersecurity, fraud and brand misuse, and corporate research and protection are especially valuable for trust and safety teams. With the right expertise in place, businesses then need to focus on empowering analysts to speed time to insight and drive better outcomes. That's where cutting edge tools for secure, anonymous research are essential.

Trust and safety teams also need the agility to overcome IT hurdles for accessing untrusted sites. It's important for analysts to tap into rich resources for gathering evidence, but all too often they are locked out (understandably) by their own administrators to avoid risk to the organization. Bad actors could also block analysts from accessing their sites from certain IP addresses.

Why trust and safety is risky business

While automated systems may flag content and usage violations as well as fraud, trust and safety analysts need to understand the whole story behind an issue. Deeper research is often needed, and that's where the risks start piling up. Analysts don't know where an investigation may lead. Scratching below the surface could require interaction with bad actors, malicious sites and untrusted resources.

Trust and safety teams could face significant risk — to the analysts, the organization and the integrity of investigations. Let's look at some common scenarios...

Acceptable use enforcement on social media

If a user or specific content is flagged as potentially abusive (whether by the content moderation system or another user), the trust and safety team must investigate. They need to reduce and prevent content abuse and criminal activity, but they also want to avoid incorrectly removing people from the platform. That means diving deeper to learn, for example, is the content really fake news or from a hate group vs. someone expressing opinions?

To get to the bottom of things, analysts might have to research off-platform, including on unsavory sites, such as forums frequented by hate groups or terrorist organizations. Anonymity is key so analyst research cannot be attributed back to their company.

Marketplace surveillance of counterfeit/stolen goods

Beyond verifying legitimate merchants, the trust and safety team for a marketplace platform needs to identify, remove (and potentially report to law enforcement) counterfeit and stolen physical or digital goods. In addition, law enforcement and counterfeit crimes units may be required to research dark web or other illegitimate marketplaces to gather evidence in order to shut down criminal activity.

Compromised accounts and fraud prevention

Online financial transactions are plagued with fraud, from identity theft on credit cards to account takeovers. Despite stringent fraud detection and prevention measures, institutions still need investigators to uncover intel behind various types of fraud.

In online communities, trust and safety analysts are tackling a rise in fake accounts used to legitimize appearances for illegal activity and spread misinformation. Often these accounts are stacked with fake likes and influencers, and sold illegally as digital assets. Researchers may need to dive beyond individual accounts and into the broader source of criminal activity.

Dealing with untrusted content and environments is risky. Investigators need to gain a complete picture for analysis, and establish a chain of evidence — without putting themselves or their company in harm's way. If analyst identities are exposed or targets get "tipped off," the time and effort of casework could be wasted. Or worse, targets could retaliate with anything from phishing, malware and DDOS attacks on company networks, to personal threats against investigators.

How isolation and managed attribution make a difference

Successful trust and safety investigations require the ability to securely isolate and anonymize browsing. Isolating investigative research away from the network eliminates risk of exposure to the company. Managed attribution enables analysts to customize their online presence to remain cloaked in anonymity. Combining these capabilities gives trust and safety teams the power to shield identities, devices and networks from risk.

Isolation

Researchers are often blocked from sites in certain categories, and need to get special access privileges for investigations. Or IT may allow view-only access on various sites, making it hard to collect evidence. Understandably, the organization needs to avoid any exposure to cyber threats, yet analysts need broader access to get the job done. Isolated, cloud-based browsing helps solve the challenges for both analysts and IT.

Isolation is a first line of defense to reduce risk. With cloud browsing and storage, investigative activity is 100% isolated from an analyst's workstation and network. Investigators can work on the same computer they use every day, yet conduct online research via a secure cloud-based service.

Isolating investigations empowers an organization to:

- **Increase access to valuable evidence:** Isolated browsing is fully protected, eliminating the need for IT to block site access or hassle with on/off privileges. Analysts gain the flexibility to research across the surface, deep and even dark web according to their organization's policies.
- **Reduce risk of exposure:** Cloud browsing safeguards the analyst's devices and network from malware, phishing, hacking and other cyberattacks.
- **Ensure a complete audit trail:** Each cloud session is unique and logged for an accurate audit trail. Trust and safety teams can more easily meet regulations and compliance, manage abuse monitoring, and maintain chain of custody, which is essential when submitting cases to law enforcement.

Anonymity with managed attribution

All too often, investigators are much more at risk than they realize. Many analysts think VPNs and private browsing are enough to cloak their identity. Not true. Browsers and websites track a user's digital fingerprint in numerous hidden ways beyond location and IP address. Online presence can be identified through browser and device attributes such as device types, OS, software/plugins installed, time zone and language settings. Add to that, a user's digital fingerprint is made more identifiable by their online behavior including search terms used, websites visited, browsing patterns, time of use, social media connections, account activity and shopping preferences.

In other words, it's tough to maintain anonymity. Yet for trust and safety analysts, it's vitally important. They don't want to tip off investigative targets, and risk them disappearing or retaliating. Analysts also need to avoid appearing to users as if they're "snooping," when maintaining due diligence over online communities and services.

Anonymity is vitally important — trust and safety analysts don't want to tip off investigative targets, and risk them disappearing or retaliating. Analysts also need to avoid appearing to users as if they're 'snooping,' when maintaining due diligence over online communities and services.

Managed attribution minimizes the risk of being tracked, identified and targeted. With this powerful capability, trust and safety investigators can increase the efficiency and effectiveness of casework, including:

- **Customize and cloak appearance:** Along with reducing risk, deception is often necessary in an investigation. Analysts can create a digital disguise that appears as if they're on any chosen device and browser. They can also assign a locational egress node that cites their origin as a certain local region, time zone and language. For example, the investigator may be in the U.S. on a Mac with Safari browser, but access a Middle Eastern marketplace and appear as if they are local and using a mobile device.
- **Eliminate geoblocking and misinformation:** Users may see dramatically different content if they visit uncloaked vs. disguised with a local IP address and browser. For example, accessing foreign sites while appearing local, analysts can often avoid propaganda and instead see content relevant to local audiences who may be considered trusted visitors.

Optimizing OSINT tradecraft

Empowering trust and safety teams to do their best work means giving them tools for best practice tradecraft. Managed attribution is a key element for investigations. But first, analysts need to understand which attributes they may need to change in their digital fingerprint. These online resources can help understand details of the target site, so analysts can adjust details accordingly:

- **StatCounter:** blend in with a local crowd using data to match the most popular browser, OS and device type used in a given region or country
- **Domain/IP Whois:** data about the registrar and ownership of a website
- **TinEye Reverse Image Search:** image recognition for content moderation and fraud detection, and tracking where images are used online
- **Social-Searcher Social Media Search Engine:** dashboard for real-time social monitoring of mentions about a company, brand, product or service
- **Social Bearing Twitter Search and Analytics:** gain insights about tweets or people based on engagement, influence, location, sentiment and more

Tools to improve investigation efficiency

Optimal tradecraft is about working efficiently, with minimal risk. Trust and safety analysts often field dozens of automated fraud and content alerts on a daily basis. The right tools for investigations can add value to the automated monitoring systems analysts use every day. For instance, teams can more quickly identify which alerts to investigate and how to get better intel, to gain deeper understanding into real threats.

Effective tools also help improve outcomes by alleviating stress on analysts. Extensive investigations into harmful or explicit content and interactions can be draining, and having more efficient workflows can reduce both strain and the likelihood of mistakes. Ultimately, the ability to conduct better investigations leads to better intelligence, which leads to better decisions.

For best practice OSINT tradecraft, look for tools that optimize four key stages of research — access, analysis, capture and audit:

Access

Reduce risk and eliminate the need for parallel infrastructure with isolated, cloud-based browsing for secure online research from any computer. Combined with managed attribution, analysts can run multiple investigations at the same time, each with unique digital fingerprints including in-region identities to avoid geoblocking. This approach to secure, anonymous browsing can be extended to any site — across the surface, deep and dark web.

> *Curious about the dark web? Trust and safety investigators are typically blocked from the most risky corners of the web, but those sites often reveal a wealth of criminal activity. It's important to understand what kinds of information can be found there, and the best tools and techniques to investigate safely. Learn more in our dark web blog series.*

Analysis

Increase efficiency with tools that build packet capture, source viewing and translation into the investigative browsing environment. In particular, with localized virtual identities (via managed attribution) and integrated translation, analysts can translate data on foreign sites without the site knowing who is actually accessing the information. It's also important to streamline multi-search workflows with automation, so analysts can run frequently used searches across multiple sites in a single click.

Capture

Simplify data collection with a browsing environment that makes it easy to capture screenshots and videos, tag assets with URLs and timestamps, and save assets into case files. And save time by automating recurring data collection with a scheduler customized for specific sites, frequency, time of day and content types. This allows the analyst to maintain tradecraft even when they're not performing the capture themselves (due to time constraints, odd hours, etc.).

Make sure collected information is securely stored off-network in the cloud, to protect the organization from malware or similar risks. Shared storage in the cloud also improves efficiency in collaboration, ensuring fellow analysts can properly access necessary information.

Audit

Keep investigations compliant with industry regulations and internal policies by ensuring all analyst web activity is logged, and logs are encrypted with organization-managed keys. Best practice tools apply usage policies to a user and their investigative cloud-based browser (rather than a device). It enables the organization to seamlessly enforce policies and log activity across all of an analyst's research on any device and any network.

Taking trust and safety to the next level

Online engagement and commerce are constantly expanding — and the challenges of trust and safety teams are evolving along with them. Investigators need faster time-to-insight and more complete information to better protect their company's online communities, services and customers.

Minimizing risk while increasing efficiency is crucial to successful results. Isolated cloud browsing, managed attribution and other tradecraft tools empower analysts to research across the web securely, anonymously and efficiently.

About Silo for Research

Silo for Research enables trust and safety teams to maintain the integrity of online platforms with faster, more secure investigations across the surface, deep and dark web. Built on Authentic8's patented, cloud-based Silo Web Isolation Platform, Silo for Research provides 100-percent protection from all web-borne threats and complete oversight of all research activity. Investigators can count on full online anonymity in an isolated browsing environment, and increase efficiency with an integrated suite of workflow productivity tools.

Trust and safety teams of the most well-known social media and digital marketplaces rely on Silo for Research. See how managed attribution can make a powerful difference for your organization — visit our Experience Center now.