Mitigating Cybersecurity Attribution Risk: An AI Approach to low collateral cyber counterattack design

Cybersecurity Glossary

Major U.S. Cybersecurity Laws

CFAA – Computer Fraud and Abuse Act (1986)

FISMA – Federal Information Security Modernization Act

HIPAA – Health Insurance Portable and Accountability Act

FERPA – Family Educational Rights and Privacy Act

CISA – Cybersecurity and Infrastructure Security Agency

CIRCIA – Cyber incident Reporting for Critical Infrastructure Act

Attribution - The process of determining who is responsible for a cyberattack.

U.S. Defend Forward Doctrine

Defend Forward - A cybersecurity strategy that seeks to identify and disrupt threats before they reach their intended target.

Hunt Forward - Operations where cybersecurity teams work within partner networks to detect adversary activity before it causes damage.

Persistent Engagement - A proactive cybersecurity approach that continuously challenges and monitors adversaries rather than reacting after attacks occur.

Common Types of Cyber Attacks

Phishing – a cyberattack where a threat actor pretends to be an employee, business, and tricks the user into giving personal credentials through email and fake links

Malware – A malicious program that steals info, damages, or controls your computer

RAT – Remote Access Trojan (Type of Malware)

Attribution Risk - The danger of incorrectly identifying the source of a cyberattack, potentially leading to an inappropriate response.

Artificial Intelligence (AI)

Technology that enables computers to analyze data, identify patterns, and assist in decision-making.

Machine Learning (ML)

A subset of AI that allows systems to learn from data and improve predictions over time.

Cyber Counter-Attack

A defensive cyber action intended to disrupt, degrade, or stop malicious cyber activity.

Low-Collateral Response

A cybersecurity response designed to minimize unintended impacts on innocent users, organizations, or systems.

Collateral Damage

Unintended harm caused to systems or individuals who were not the intended target of a cyber operation..

Denial of Service (DoS) Attack

An attack that attempts to make a system or service unavailable by overwhelming it with traffic or requests.

Distributed Denial of Service (DDoS) Attack

A DoS attack launched simultaneously from many compromised devices.

ICMP Flood

A DoS attack that overwhelms a target with excessive ICMP (ping) requests.

SYN Flood

A DoS attack that exploits the TCP connection process by sending incomplete connection requests.

UDP Flood

A DoS attack that overwhelms a target with large amounts of UDP traffic.

HTTP Flood

An application-layer attack that overwhelms a web server with excessive HTTP requests.

Connection Hold Flood

An attack that keeps many network connections open, consuming server resources over time.

SMB/RPC Flood

An attack targeting Windows networking services such as file sharing and remote procedure calls.

Attribution Confidence Score

A measure of how certain analysts are that a specific actor is responsible for a cyberattack.

Decision-Support Tool

A system that assists analysts in evaluating cyber incidents and selecting appropriate responses.

Automation Bias

The tendency for humans to overtrust automated systems and recommendations.

Explainable AI (XAI)

AI systems that provide understandable explanations for how decisions are made.

Blast Radius

The estimated scope of systems, users, or organizations that may be affected by a cyber event or response.

Human-in-the-Loop

A decision-making approach where humans maintain final authority while AI provides recommendations.