Protecting Privacy in Algorithmic Analytics Programs

Briefing for the Information Security and Privacy Advisory Board

Dan Chenok
Chair, DPIAC Cyber Subcommittee
Former Chair, ISPAB
March 24, 2016

Background

Summary:

- DHS (NPPD) has a pilot program underway to develop techniques and assess effectiveness of “Algorithmic Analytics” (AA)
  - The real-time analysis of traffic, seeking to identify anomalies that could point to malicious activity
- Privacy Office and NPPD worked together on potential privacy issues raised by these pilot programs
  - Chief Privacy Officer requested that the DPIAC, through the Cyber Subcommittee, make recommendations on how best to protect privacy in such programs
  - Cyber Subcommittee met with NPPD and CPO staff for information briefings, in process of developing recommendations for DPIAC presentation and release

Approach

1. General Considerations, defined scope of inquiry

2. Addressed key considerations affecting program

3. Identified Potential Privacy Protections at each program stage (Collection, Use, Sharing, Retention, Access, Disposition)

*Built on related findings from prior DPIAC reports

General Considerations

- "Behavioral Analytics" as a term? Subcommittee concerned that this implied tracking individuals
- DHS briefings and discussions led to recommendation for alternate term, such as "Algorithmic Analytics"
  - Descriptive of actual activity: empirical analysis of network traffic activity, collected using automated means, seeking to identify anomalies that could point to malicious activity
    - Establish baselines for patterns of traffic, use machine algorithms to spot anomalous patterns from common baseline
    - Ability to determine potential malicious traffic without knowing a predetermined signature
    - Analyst then does further review to determine if anomaly is associated with a potential problem event (e.g., vulnerability, threat, or incident)
  - Is not signature-based assessment of specific computer-device or application traffic, or assessment of human behavior

General Considerations

- Commercial companies engage in AA today – technologies used include (provide potential benchmarks):
  - Front-end authentication
  - Transaction Monitoring
  - Real-Time Queries
  - Risk scoring matrices
  - Egress tracking
- DHS Pilot: Logical Response Aperture (LRA)
  - Follows traffic entering and exiting system
  - Strong privacy protections
  - Can be model for privacy protections in expanding AA, from agency to government to industry partners
- Mobile devices present special issues and considerations for PII

Key considerations

- How is PII affected by AA?
  - Likely to be minimal, situations include:
    - If AA points to individual accounts, could be mishandled
    - AA data could be correlated with PII during analysis phase
    - FIPPs come into play
  - When PII involved, rely on existing policies for protection
  - Three categories:
    - All traffic, no need to retain and basic protections apply
    - Anomalies point to malware, further investigation requires special protection for PII
    - Sample data for training, need to strip PII

- Data Quality and Integrity key throughout
  - Includes content of information as well as metadata
  - Focus on data integrity throughout, AA data can be a target
  - Address data governance, including records management

Key considerations

- Accountability
  - Human oversight is vital
    - Allow ongoing reviews, redress
    - Individual info needs protection when shared or analyzed
  - Algorithms should be accessible to human review at periodic intervals, with exceptions
  - More than one reviewer should be engaged
  - Focus on sound procedures, appropriate design, fairness in selecting targets, treatment of PII
  - Privacy Office review of process

Potential Privacy Protections at Each Stage

- Collection
  - “Strip and Encrypt”, de-identify (where feasible)
  - In transit and at rest
  - Provide notice
  - Transparent criteria for what to collect
- Use
  - Define multiple uses – network protection, fraud, website management, criminal acts, etc.
  - Establish process for each use, and use limitations
  - Protocols for follow-up analysis
- Sharing
  - How to interface with sharing centers (CERT, NCCIC, etc.)?
  - Technical sharing parameters need further focus
  - Define rules for sharing with law enforcement, private sector, ISACs/ISAOs, etc.

Potential Privacy Protections at Each Stage

- Access
  - Link with prior DPIAC recommendations
  - Limit personnel who can work with data
  - Determine rules for when/how to access
  - Develop controls for access, including logs

- Retention and Disposition
  - Consider how long to retain
    - Limit to shortest time needed for program purpose
    - Caution in establishing fixed time frames – need flexibility as technology and uses evolve, but make timelines transparent
  - Separate AA data from analytics on that data
  - Establish disposition protocols – how ensure all copies and versions are addressed?
  - Need for security throughout lifecycle and to disposition, consistent with government rules
  - Consider audits, periodic reviews