Discussion Draft

Straw man architecture for International data exchange and collaborative analysis

John C. Mallery (jcma@mit.edu)
Computer Science & Artificial Intelligence Laboratory
Massachusetts Institute of Technology

7/8/2011 3:27:41 PM

Presentation at the "BIC Session on Building a Long-term INCO Strategy in Trustworthy ICT," First SysSec Workshop, Vrije Universiteit, Amsterdam, July 6, 2011.

INCO-Trust Workshop on International Data Exchange with Security and Privacy, New York, May 5, 2010

- **Claim:** International collaboration and coordination can reduce defensive gaps across the OECD and build crisis-response capacity.
- **Leverage:** Bias work factors in favor of defense and against cyber attack
- **Approach:** Exchange data related to cyber crime, attack patterns and best defense practices
- **Research:** Motivate research via technical needs of the data sharing scenario
  - Bring together researchers in security and privacy
  - Focus on relevant research challenges
  - Leverage national concentrations of expertise

Inco-TRUST-2010: Technical Research Challenges

- Representation and structure of data
  - Including standards for structured text
- Policy representation and understanding
  - Law and social norms
  - Expressive policy languages
  - Policy analysis
  - Conflict detection
  - Formalization and semi-automated enforcement
  - Usability)
- Architectures and enforcement
  - Cryptography
  - Private data analysis
  - Watermarking
  - Identifying anomalous behaviour
  - Audit and accountability
  - Managing risks and economic analysis
  - Access control and other preventive techniques
  - Issues regarding integrity of data, undo decisions, provenance, retraction, and update
  - Process-centric vs. Data-centric architectures
  - Compositional enforcement of policies under constraints
  - Conforming to legal requirements and social norms
- Development platform and test bed
  - Interface for contributing data
    - Network attack data, DHS, clinical studies data, vulnerability data
  - Cloud computing distributed across administrative zones
    - Drives research on changing policies and associated enforcement
    - Specialize to different domains, like healthcare, DRM

Benefits Of International Cyber Data Sharing

- Data sharing
  - **Trends:** Retrodictive cyber statistics across the OECD
  - **Anti-crime measures:** Cyber crime targets, vectors, methods, counter-measures
  - **Closing defensive gaps:** comparison of defensive coordination and best practices
  - **IP Protection:** Detection and prevention of industrial espionage
- Expertise integration
  - Focus collective expertise on important cyber data and analysis tasks
  - Faster learning and transfer into operational practice
- Collaboration and coordination
  - Preventing replay of attacks across countries and sectors
  - Reducing defensive gaps across the OECD
  - Crisis response
- Research and development coordination
  - Leverage and combine national expertise

Data Collection and Sharing Goals

- Build shared awareness and understanding of cyber phenomena across countries
  - Employ shared data collection methodologies
  - Integrate measurements of phenomena across borders
  - Focus early on cyber crime and cyber economics
- Create comparable transnational data sets
  - Capture cyber breaches, attack patterns, best practices, defensive coordination
  - Include aggregate data on crime, black markets, economics, state-state interactions, long-term cyber-fueled transformations
- Field a cyber data sharing framework that helps countries to:
  - Collect cyber data for compatible sharing
  - Fuse data to create common situational awareness
  - Manage national legal impediments to sharing via derived or aggregate data or by recommending harmonization steps
  - Exchange derived data in real time
  - Provide mechanisms for controlled drill down needed for law enforcement, advanced persistent threats (APT) or cyber emergencies
- Develop shared collection, fusion, analysis, and response capabilities

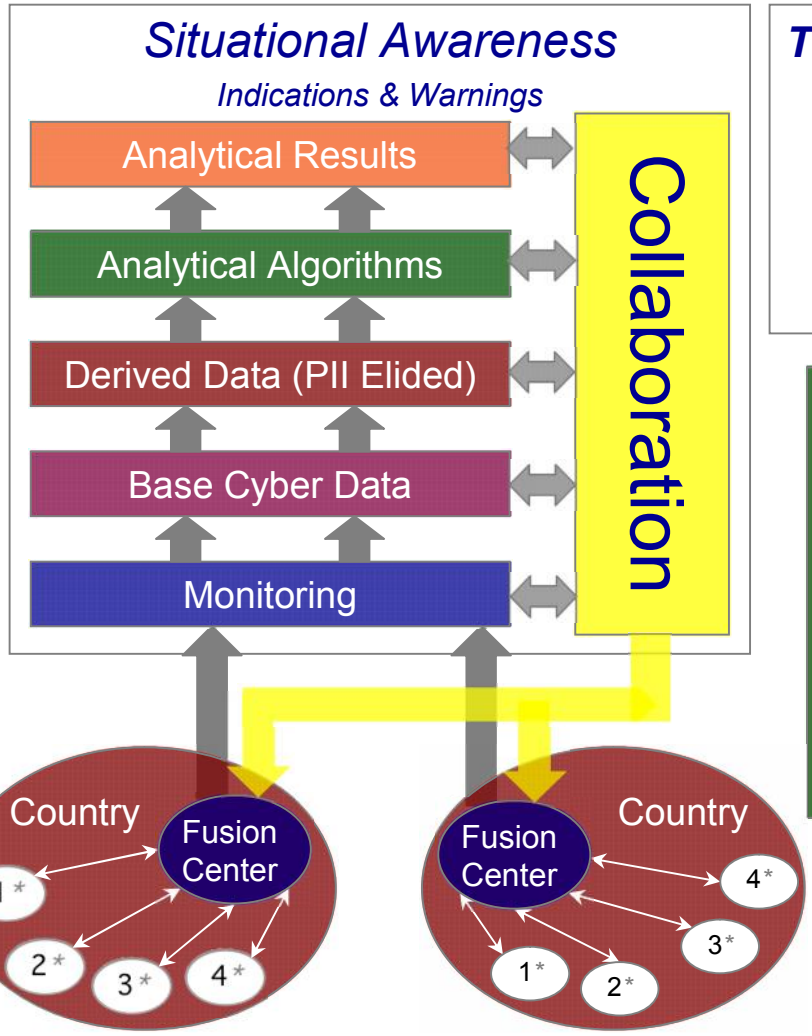
Precedents Can Inform The Architecture

- European Network & Information Security Agency
  - Collects, analyzes, disseminates data on InfoSec in pan European context
- DHS Predict (US)
  - Legal framework for sharing cyber data with US
  - International framework in progress
- Wombat Project
  - Collaborative sensors for Internet malware and attack data
- European Public-private Partnership For Resilience
  - Critical information infrastructure protection
- Financial Services Information Sharing and Analysis Center (FS-ISAC)
  - Organizations submit information anonymously
  - Data received by members cannot be attributed to any specific organization
  - ISAC was based on NYC model
  - ISAC was created by financial services industry
- National Cyber Forensics and Training Alliance (US)
  - Non-profit that integrates information and analysis for the financial services sector across private, public and academic communities
- Confickr Working Group
- Phishing Sharing
  - Anti-Phishing Working Group
  - Digital Phishnet
- Others?

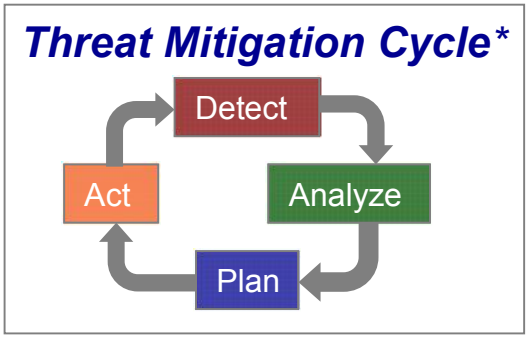
Political and Legal Challenges

- Political and legal barriers
  - Divergent legal requirements
  - Government procedures for handling classified information
  - Export controls
  - Proprietary data
  - Privacy
- How can progress be made quickly before legal and regulatory harmonization is addressed? (10 years or more?)
  - Share derived data
  - Share patterns to be detected in 1st order data
- Create country-level fusion centers (CERTS?)
  - Provides governments with control over national data and analysis
  - Manage drill down for exceptional cases

International Cyber Data Sharing Architecture

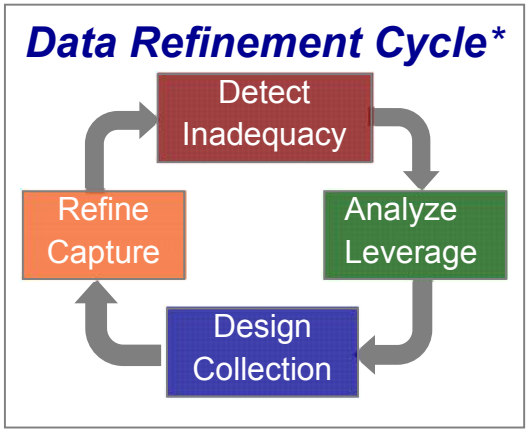


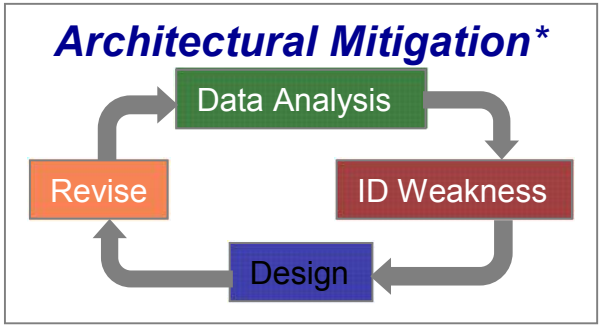
* Sector level

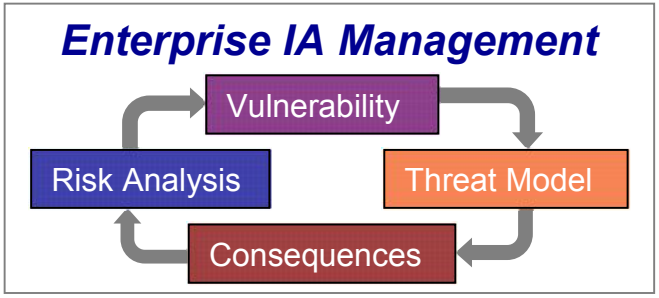


Data Types

- ✓Statistical data
- ✓Attack vectors
- ✓Signatures
- ✓Algorithms
- ✓Remediations
- ✓Targets
- ✓Threat Intel
- ✓Best practices
- ✓Case studies







Key Idea: Data Generality vs. Specificity

- Data sensitivity often correlated with specificity
  - Sharing easier: Aggregate data
  - Sharing harder: Specific, identifiable data
- Three Legal frameworks
  - Aggregate data suitable for national accounts statistics
  - Intermediate data provides more structure but no revelation of identities or “private” data
  - Specific data gives full details with PPI obfuscated or not
- Specific data will likely require a PREDICT like legal framework
  - Establish framework for provider-consumer specific agreements for shared data
  - Provide special handling procedures for sensitive data
- Incremental access via general -> specific can reveal whether access is needed
  - Fine-grained security can support precise access
  - New access control model based on abstraction?

Defensive Complexity Analysis: Coordination Reduces Search Space for Defenders

- Attacker search leverage
  - Integrated organization
  - Focus on target
  - Choice of attack vector(s)
  - Selection of place and time of attack
  - Black markets for crime ware
- Defender search leverage
  - Shared situational awareness (data, info)
  - Constrain search by linking data across dimensions (attack vectors, value at risk)
  - Shared detection
  - Shared responses
  - Shared best practices
  - Sharing expertise
  - Shared analytical tools and methodologies
  - Focused and scalable collaboration
  - Shared R&D
- Amortize effort to establish frameworks for international data sharing
  - Base: Data, information, knowledge, algorithms
  - Expertise: Index expertise around data topics
    - Better ability to understand and use the data
    - Enable focused collaboration
  - Collaboration: Refine the data, practices and responses
    - Sectoral practitioners, stakeholders
    - Cross-sectoral synergies
    - R&D to improve data, practices and responses
  - Architecture: Evolve the sharing and collaboration system
    - Task-driven legal and regulatory harmonization
    - Security & collaboration research

Incentives

- General
  - Data views into cyber only possible through transnational integration
    - Retrodictive data to understand trends
    - Predictive data for early warning
    - Identification of best practices
    - Crisis response mode
  - If you contribute, you get the benefits of other's contributions.
    - Public goods and free rider challenges
  - Slice the data vertically so that participants can subscribe to those verticals for which they collect and submit derivatives
- Governments
  - Control of national fusion centers
  - Operate under national laws
- Security Industry
  - More collection by other parties enables more value-added products & services
- Private Sector
  - Better cyber data on which to base decisions, more accurate risk assessment across sectors
  - Early warning as crime models developed and applied to sectors
- Technical reinforcement of incentives for participants
  - Resilient mechanism design
  - Interdependent risk models to inform design
  - Game theory

Caveat: Cyber defense data can be relevant to offense

- Duality of defensive and offensive information
  - Understanding defensibility of targets can inform attacks against them
  - Wide scope and comprehensiveness of cyber data create an irresistible target for insiders and well-resourced outsiders
- Certain data must be closely controlled
  - Weaponizable cyber data
    - Control data for botnets
    - Malware
    - Vulnerability library
    - Critical sector defense practices
  - Data with high monetization potential
    - Effective attack patterns against valuable targets

Implications of Data Relevant to Cyber Offense

- Access control and accountable use are critically important
  - Segregate dangerous data
  - Require high assurance handling
- Trustworthiness of data sharing architecture and implementation is crucial for buy-in
  - Design for high attacker work factor
  - Protect against insiders
  - Partition and differentially encrypt data
  - Combine on an as-needed basis under strong assurance
  - Produce less dangerous derived products
- Sharing will be as deep or as shallow as trust among cooperating countries
  - Calibration of intent
  - Ability to enforce access control policies

Data Harmonization

- Collaborative processes required for developing:
  - Shared vocabularies for describing cyber data and phenomena
    - Cross human language mappings
  - Domain ontologies data
    - Standardized data formats within domains
    - Approach for data format evolution
    - Cross-ontology linkages
    - Provenance metadata
    - Security policies – appropriate use
  - Derived data definitions
    - Algorithms for computing derivations
    - Metadata for outputs
- Approach
  - Start from existing datasets
  - Work towards integration of base datasets
  - Define generalization planes on datasets
  - Evolve datasets based on experience

Template for Recommended Data Collection

- **Question:** What is the scientific or policy-relevant question?
- **Data Requirements:** What data would enable us to answer it?
- **Analytical Tools:** What kind of analysis techniques are relevant?
- **Existing Data:** If data already exists,
  - Who controls it?
  - How can it be accessed?
- **New Data:** If new data is required,
  - Who can collect it?
  - What practical or legal issues are involved?
- **Data Comparability:** How can this data be compared?
- **Indicators:** What indicator(s) can be devised from this data?
- **Collection Difficulty:** How feasible is it for someone to collect and share this data?
  - What are the legal barriers?
  - What are the trust or privacy constraints?
- **Collection Priority:** How important is this question and the data necessary to answer it?

Key Questions

- **Data:** What cyber data should be shared?
  - What domain?
  - What purpose?
- **Synergies:** What synergies arise from integrating data across national boundaries?
- **Impact:** How will it help participating countries?
- **Incentives:** What are the incentives for providing data?
- **Quality:** How can the integrity and quality of data be assured?
- **Availability:** How can data be made available in useful formats and in time to be relevant?
- **Risk:** How should data sharing risks be managed?
  - What risks are involved in assembling and sharing data?
  - How can data be sliced or aggregated to reduce risks?
  - How can access be controlled with incremental revelation to reduce risks while enabling benefits?

Conclusions

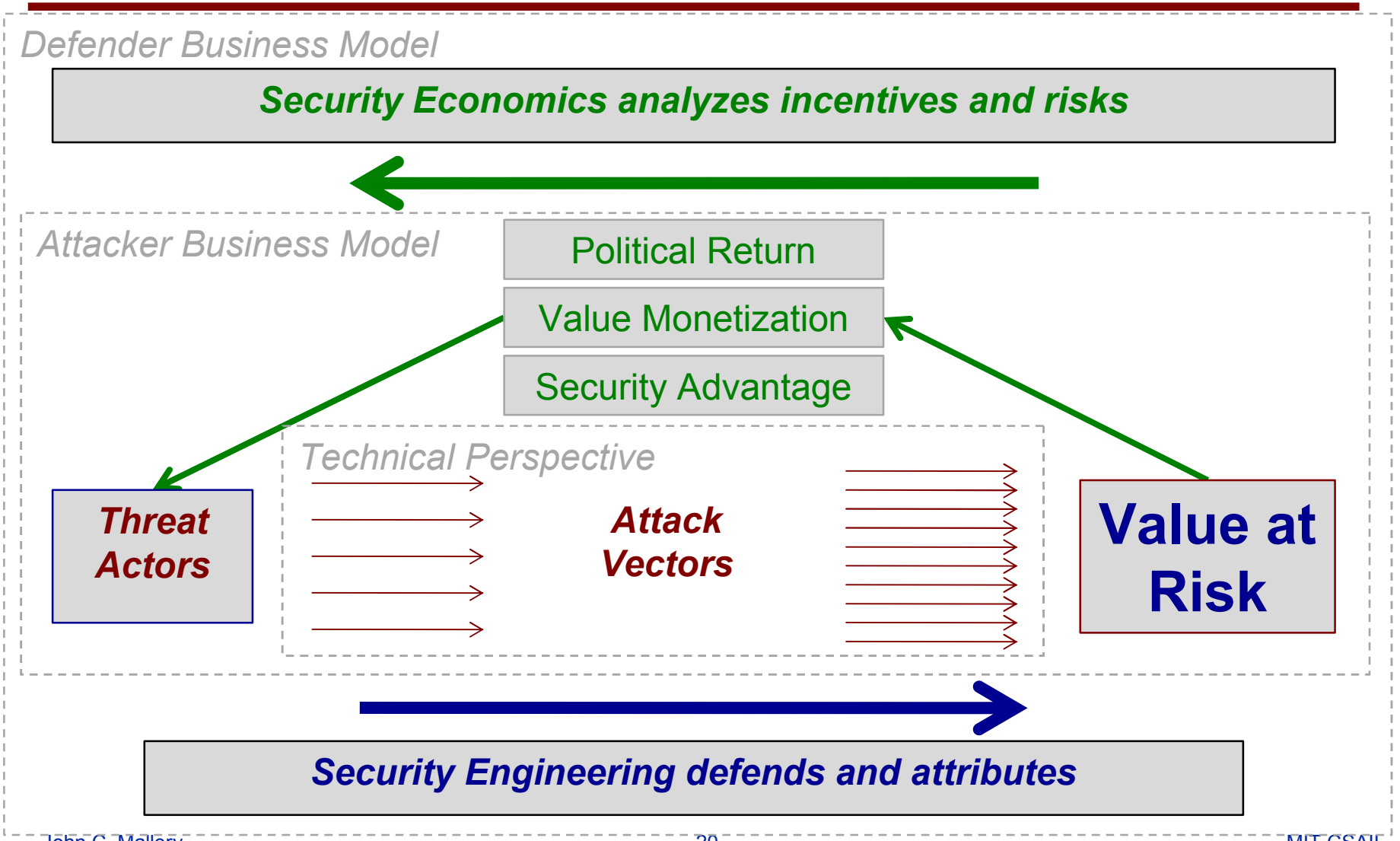
- Start with a narrow yet useful data dimensions
  - Begin with lower sensitivity data
  - Focus on data to characterize aggregates
  - Easy -> hard sharing
  - Defer dangerous data
- Decompose data by sensitivity horizontally and vertically
- Build warning data around communities of expertise able to understand and analyze data
  - Specific sectors
- Target applications to support:
  - Situational awareness
  - Defensive coordination
  - Defeating attacks (or increasing uncertainty of success)

Appendix

Threat Actors And Capabilities

Threat Actors	Motive	Targets	Means	Resources
Nation States During War Time	Political	Military, intelligence, infrastructure, espionage, reconnaissance, influence operations, world orders	Intelligence, military, broad private sector	Fully mobilized, multi-spectrum
Nation States During Peace Time	Political	Espionage, reconnaissance, influence operations, world orders	Intelligence, military, leverages criminal enterprises or black markets	High, multi-spectrum, variable skill sets below major cyber powers
Terrorists, Insurgents	Political	Infrastructure, extortion	Leverage black markets?	Limited, low expertise
Political Activists or Parties	Political	Political outcomes	Outsourcing?	Limited, low expertise
Black Markets For Cyber Crime	Financial	Hijacked resources, fraud, theft, IP theft, illicit content, scams, crime for hire	Tools, exploits, platforms, data, expertise, planning	Mobilizes cyber crime networks
Criminal Enterprises	Financial		Reconnaissance, planning, diverse expertise	Professional, low end multi-spectrum, leverage of black markets
Small Scale Criminals	Financial		Leverages black markets	Low, mostly reliant on black markets
Rogue Enterprises	Financial	IP theft, influence on sectoral issues	Outsourcing to criminal enterprises?	Sectoral expertise, funding, organization

Integration of Technical and Economic Perspectives



Architecture

- Establish national cyber data fusion centers
  - Collect shared base data across public and private sectors
  - Apply shared algorithms for aggregation, pattern matching and creation of derived data
  - Share the derived data
  - Build redundancy into the collection mechanism to enable checking of data validity
- Link these centers together for sharing
  - Provide methods to assure the integrity of data
  - Provide appropriate access control policies
- Make data available to appropriate 3rd parties to perform value-added analysis
- Accept feedback from all participants on improvements to data collection and analysis
  - Track successes, near misses
- Tight integration with key sectors
  - Telcos, ISPs, network & computing infrastructure (implements ICT)
  - Critical infrastructures
  - Financial sector
  - Major online application infrastructure
  - IP-based sectors (industrial espionage)
  - E-business (readily accessed via Internet)

Recursive Architecture

- Create a scalable sharing, fusion and collaboration architecture
- Apply the architecture the the following levels:
  - Sectors
  - Countries
  - Internationally
- Effectiveness will be as good as the engineering and security architecture investment in the systems
  - Trust in the platform will govern how much data is shared
- Field higher standard systems at all levels
- Amortize the cost at multiple levels

Real-time Forensics

- Real-time forensics may mitigate the need for long-term storage
  - RTF could trigger the logging tool proactively
  - But statistical RTF probably requires initial training on existing data
- Need data for attribution in the short term, but also need data for simulation (Monte Carlo) to look for patterns
  - Entity analytics for anomaly detection
- Zero-day AI-based anomaly detector
  - Symbolic RTF may be possible with good theory
- Positive detection of deviation from modeled functions

Data Considerations

- Comparable characterization of cyber crime across countries by:
  - Sectors targeted
  - Methods of attack & coordination
  - Vulnerabilities exploited (technical and organizational)
  - Nature of criminal organizations (including black markets)
  - Precursor signatures
  - Detection signatures
  - Effectiveness of defensive coordination
  - Countermeasures
- Capture of sufficient time window data to detect APT footprints
  - Retention of ISP and traffic logs
  - Enterprise network sensor data
  - Compression via known patterns
- Layer data by generality
  - Aggregates easier to collect and manage
  - Specifics better for response to criminal or state activities
- Support real-time response to cyber crime or other attacks
  - Higher assurance handling required
  - Need based access
- Measure contribution of data to anti-crime efforts
  - Feedback to improve data and reinforce cooperation

Data Planes

- Technology Plane
  - Focus on the range of technical vulnerabilities and technical approaches to defense
  - Includes new attack surfaces like cloud computing and mobility, but also improving defensive technologies
- Cyber Crime and Criminal Justice
  - Focus on cyber crime motivated by financial gain, prevention, detection and prosecution
  - Includes economics of cyber crime, forensics, industrial espionage, vigilante activities, international cooperation
- Economic Plane
  - Focus is on data supporting policy moves to improve market response to cyber security
  - Includes industrial organization in the IT capital goods sector, risk management, actuarial data and insurance and analysis of potential government intervention
- Defensive Coordination
  - Focus on sharing of threat, vulnerability, breach and response data as well as best practices within sectors
- State-centric Cyber Interactions
  - Focus is on interstate cyber espionage, sabotage, preparation of the battlefield and cyber attacks
  - Includes indications & warnings, cooperative defense, norm development
- Cyber-fueled Long-term Transformations
  - Focus is on the transformations within modern economies and international systems arising from ubiquitous integration of computation and global networking
  - Includes changing action possibilities for old and new groups, whether economic, political or affinity.

Tension Between Openness & Sensitivity

- Low sensitivity information can be shared to enable interesting applications or analyses
- Open cyber data empowerment
  - Leverage: Enable 3rd parties to develop value-added analyses
  - Evolve: Feedback ideas for improvement into data requirements and to providers
- Highly sensitive information cannot be widely shared
- Need to strike a balance between openness and safety
- Provide appropriate data handling regimes according to safety criteria

Harmonized Retention Requirements

- Data retention policies should match the statute of limitations
- Law enforcement seems to like 3 years
- But, APT is longer term, so longer retention will help with APT

Business Process Changes To Support Sharing

- What business processes are needed to support data sharing?
- Who pays the cost of data collection & sharing?
- What incentives will motivate senior management?
- What data control guarantees are needed enable sharing?
- What security and privacy technologies are necessary to enforce sharing policies?

Conflict Management

- Disagreement will exist and orderly mechanisms for principled resolution will be required
- Domains
  - Definitions
  - Data categories
  - Data formats
  - Standards
  - Legal & regulatory regimes
  - Purposes & political constraints
- Security
  - Data security policies & enforcement
  - Credentialing, authorization and authentication
  - Host, network and crypto assurance levels

Enabling Technologies

- Core
  - Harmonized data collection strategies
  - Tools for collection, storage and pattern matching
  - Patterns of criminal and APT behavior
  - Data interoperation standards (e.g., semantic data web)
  - Binding of legal and regulatory requirements to process and security architectures
  - Usability
- Security
  - Identifying anomalous behavior
  - Audit and accountability
  - Managing risks (economic analysis)
  - Access control, preventive
  - Approaches to access control across administrative boundaries
  - Compositional enforcement of policies under constraints
  - Remote policy enforcement
  - Integrity of data
  - Undo decisions, provenance, retraction, update
  - Sandboxed computation on sensitive data
- Cryptographic techniques
  - Data splitting
  - Differential privacy
  - Checking the integrity and provenance outputs from computations on the data
  - Private data analysis
  - Watermarking (aggregate information)
  - Arithmetic (+, x) helps elicit data without divulging identity
    - Inputs are encrypted
    - Operation performed on cyber text
    - Outputs decrypted using keys that cannot decrypt the inputs
    - Extra credit: audit system to check the data integrity and reliability

Supporting Technologies

- High integrity storage with fine-grained security
- Pattern-based compression techniques
- Techniques for obfuscation of private information, including identity
- Policy representation and understanding
  - Understanding the law, social norms (higher level concepts, meta-language?)
  - Expressive policy languages
  - Policy analysis and conflict detection
  - Formalization and semi-automated enforcement
  - Usability
- Revocable anonymity
- Technical reinforcement of incentives
  - Resilient mechanism design
  - Interdependent risk
  - Game theory

References

- AICPA, “Trust Services Principles and Criteria,”
  - Accounting association data protection principles
  - http://infotech.aicpa.org/NR/rdonlyres/0D6000E3-9F5F-4CCF-9D14-3B68343AB93C/0/FINAL_Trust_services_PC_Only_0609.pdf
- APEC, “Privacy Framework,”
  - http://www.ag.gov.au/www/agd/rwpattach.nsf/VAP/(03995EABC73F94816C2AF4AA2645824B)~APEC+Privacy+Framework.pdf/$file/APEC+Privacy+Framework.pdf
- DHS Predict
  - https://www.predict.org/
- European Network & Information Security Agency
  - http://www.enisa.europa.eu/
- European Public-private Partnership For Resilience
  - http://ec.europa.eu/information_society/policy/nis/strategy/activities/ciip/impl_activities/index_en.htm
- Financial Services Information Sharing and Analysis Center (FS-ISAC)
  - http://www.fsisac.com/
- National Cyber Forensics and Training Alliance
  - http://www.ncfta.net/
- OECD, “Privacy and Personal Data Protection,”
  - http://www.oecd.org/dataoecd/30/32/37626097.pdf
- Symantec Wombat Project
  - http://www.wombat-project.eu/