Data Logistics Requirements for eCrime Event Data Sharing

Peter Cassidy

Secretary General – APWG

www.antiphishing.org

pcassidy@antiphishing.org

Electronic Crime is Different

- Forensic narrative in eCrime is most often elusive:
  - Never as easy as ‘guy robbed a bank and fled on foot, south on Main St.’
- Data voluminous and largely redundant
  - Human processing made impossible by overwhelming volume of data and disparate file formats
- Data scattered across disparate jurisdictions and venues
  - Disparate data protection laws complicates collection and sharing of eCrime data
  - Industry, holding larger proportion of forensically potent data are left under a cloud and exchange more ad hoc than formally

Data Sharing and Cooperation is Key to Success of LE Effort

- Cooperation between national law enforcement agencies is growing and resulting in successful prosecutions
  - Recent US/Egypt Phish Phry Busts
  - Multinational LE efforts growing in scope and success
- CoE Cybercrime convention
- What's needed to make ecrime law enforcement investigation as fast as the crimes themselves?

Automated Machine Processing of eCrime Data

- Too much data for human handling
- Data shifting in relevance and context too quickly for human tracking
- Many types of data that contain clues about ecrimes not easily human-readable
- Solution: systematized and automated processing of ecrime event data
- Where do we start?

Cybercrime LE Cooperation Needs Formalized Data Logistics

- Keystone and Foundation: Common terminal file format
- eCrime reporters have a consistent schema to use when issuing reports
- eCrime response and investigative correspondents can read each other’s data
- eCrime response correspondents’ machines can process each other’s data according to systemized routines
- Appropriate vessel to house the disparate types of data relevant to an ecrime and make them accessible to investigators and their machines:
    - Text-based data
    - Executables such as crimeware keyloggers

IODEF Extensions XML Schema for eCrime Reporting

Extensions to the "IODEF-Document Class for Phishing, Fraud, and Other Crimeware" proffers:

- Structured data model allows forensic searches and investigations to be automated/scripted with ease using a standard XML schema
  - Multiple language capability
  - Non-ambiguous time-stamps
  - Reports are human-readable in any XML-capable browser
  - Multiple parties – brandholders; security professionals, CERT personnel and LE - can add to a report and build the story
  - Purpose-built for ecrime
  - Extensible to adapt to new ecrimes

```
+--------------------+
|  PhraudReport      |
+--------------------+
| STRING Version     | <>--{0..1}--[ PhishNameRef ]
| ENUM FraudType     | <>--{0..1}--[ PhishNameLocalRef ]
|                    | <>--{0..1}--[ FraudParameter ]
|                    | <>--{0..*}--[ FraudedBrandName ]
|                    | <>--{1..*}--[ LureSource ]
|                    | <>--{1..*}--[ OriginatingSensor ]
|                    | <>--{0..1}--[ EmailRecord ]
|                    | <>--{0..*}--[ DCSite ]
|                    | <>--{0..*}--[ TakeDownInfo ]
|                    | <>--{0..*}--[ ArchivedData ]
|                    | <>--{0..*}--[ RelatedData ]
|                    | <>--{0..*}--[ CorrelatedData ]
|                    | <>--{0..1}--[ PRComments ]
+--------------------+
```

•Data fields can be selectively encrypted to protect data from viewing by parties who are not part of, for example, a data protection convention

•Purpose built nature gives it unique relevance for eCrime event reporting

IODEF Extensions XML Schema for eCrime Data Elements

- Data elements common to phishing, fraud, and other ecrime allows the reporter to specify elements of an event:
  - The fraud source and target of crime, such as a bank
  - The Web servers involved
  - Copy of the crimeware used in a specific e-crime event with a unique digital fingerprint
  - Domain Name Service (DNS) and registry information
  - Evidentiary files of a website's content
  - Pointers to other, related archival data resources
- Extensible to adapt to forms of ecrime as criminal expertise and ambition evolves

eCrime Schema Provenance

- In 2003, APWG began clearing ecrime event data for members
- Basic schema reports URLs of phishing attacks with limited application
- URL Block List clears up to 50,000 discrete URLs for APWG member companies each month
- Limited data: enough for advising consumers and tipping off security teams
- Still not enough for forensic applications
- Members asked for a number of additions
- All are represented in this XML schema

Mine of eCrime Data Waiting for Common Schema for Processing

- Phishing attack URLs and related attack data
- Botnet IP addresses and related attack data
- Botnet command and control addresses
- Binaries of crimeware, phishing kits and botnet propogation programs
- Malware URLs
- Spam centers and operators
- Malevolently registered domains and related WHOIS data
- Registries and records of malevolently registered domains in their TLD
- IP block space/ASN data
- Human intelligence
  - Vacation photos on personal websites

Applications Enabled by an eCrime Reporting Format

- Enterprises (e.g. a group of banks)can quickly consolidate ecrime report databases to present a case to law enforcement
- Private security firms can share data quickly to indentify and characterize gangs which are causing losses to their client companies
- National CERTs, coordinating investigations into phishing attacks, can combine ecrime event databases to find corresponding data points in attacks launched in one country against targets in another
- Public sector law enforcement agencies and private enterprises can combine ecrime event databases to analyze for trends and clues to inform case initialization
- Public sector law enforcement agencies can quickly assemble relevant ecrime event data around a formerly unidentified suspect whose identity has been surmised as being party to known crimes
- All parties to development of an existing law enforcement case can program their systems to automatically direct reports of pre-determined characteristics to the appropriate investigators

So Is All This Data Exchanging Really ‘Law Enforcement’?

- Only when it is used for case formation – far rarer event that when used for animating security protocols
- Increasingly, the eCrime forensic databases developed by APWG and others appear closer to public health data exchange than a law enforcement mutual aid agreement
- So maybe worthwhile seeing how far the analogy can extend

Clearinghouses for Disease Data A Long History in Public Health

- Conference of Venice in 1892 set up protection from cholera transported through the Suez Canal
- l'Office International d'Hygiene Publique in early 1900s, formed from the preceding Conference Sanitaires Internationale
  - Established data exchange of disease data
  - Organized to protect signatory nations from diseases borne by maritime trade – the Internet of its day
- Their legacy is the World Health Organization and its formal protocols for health data exchange

Can eCrime Data Exchange Assume a Public Health Model?

- Maybe
- Large barriers to formal data exchange and usage protocols employed by public health agencies
- Real and apparent conflicts between data protection laws and ecrime data exchange need to be resolved
- Until then, ecrime data exchange – especially from private sector where most data is collected – will remain casual and episodic

For Further Information Contact:

Peter Cassidy

pcassidy@antiphishing.org

+1 617 669 1123