ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

IKEA
Departement Informationstechnologie
und Elektrotechnik

An Architectural Approach to Inter-domain Measurement Data Sharing

Brian Trammell
Communication Systems Group, ETH Zürich
DUST 2012, San Diego, 15 May 2012

Questions and Wishes

- common platform for darkspace analysis
  - “let’s use the same software to make it easier to share data”
- common interface for collaborative interface
  - “what sort of interface do we provide to a running window of data?”
- analyze unsolicited traffic on lit space
  - keep doing “darkspace”-like studies in an v4-scarce world
  - but, assumptions about privacy don’t hold at all anymore.
  - even “one-way” traffic isn’t all unsolicited

Privacy issues in measurement data sharing

- In the general case, most collected data simultaneously quite sensitive and completely uninteresting.
- Darkspace collection suffers less from these concerns:
  - no “legitimate” user traffic, so no “users” to threaten; however:
  - illegitimate user traffic (compromised host IPs, occasional payload)
  - structure of the darkspace itself
- Anonymization alone not a solution to the problem
  - any attacker that can induce or otherwise know details about traffic can reverse identifiers with ~24 bits of information per

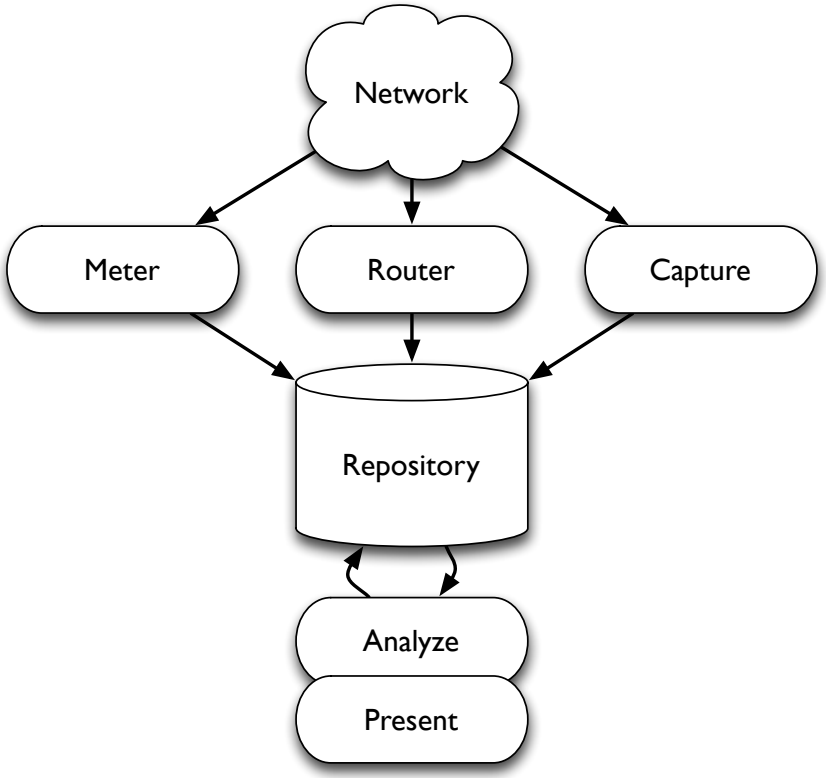
    (Burkhart et al., “The Role of Network Trace Anonymization Under Attack, ACM CCR Jan ‘10)
- Any solution must include policy framework
  - Semi-open consortia of operators/CSIRTs/researchers/etc.

Scalability issues in measurement data sharing

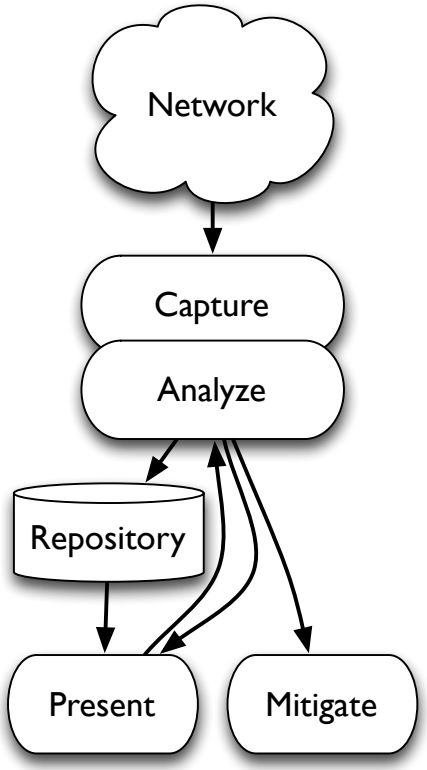
- Passive network measurement collects [insert impressive word for “big”] amounts of data.
- Separation of collection and analysis presents scalability issues at the sharing interface.
- Data volume in darkspace directly proportional to size of observation surface; bigger surfaces are more useful.
- Usual methods of addressing this:
  - limit retrospection interval (“last two weeks”)
  - limit fidelity for older data (sampling, aggregation)
  - spend a lot of money on disks…
  - …and don’t underestimate the bandwidth of a delivery truck.

The DEMONS Approach

Centralized

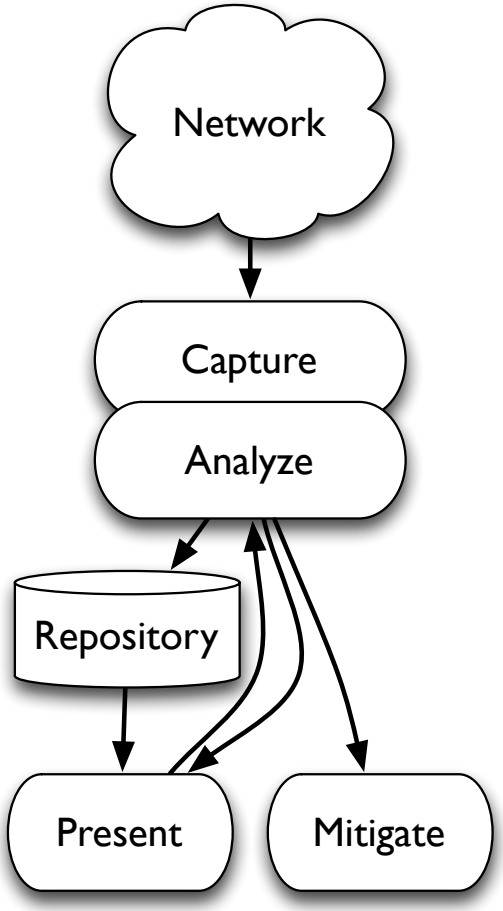


DEMONS



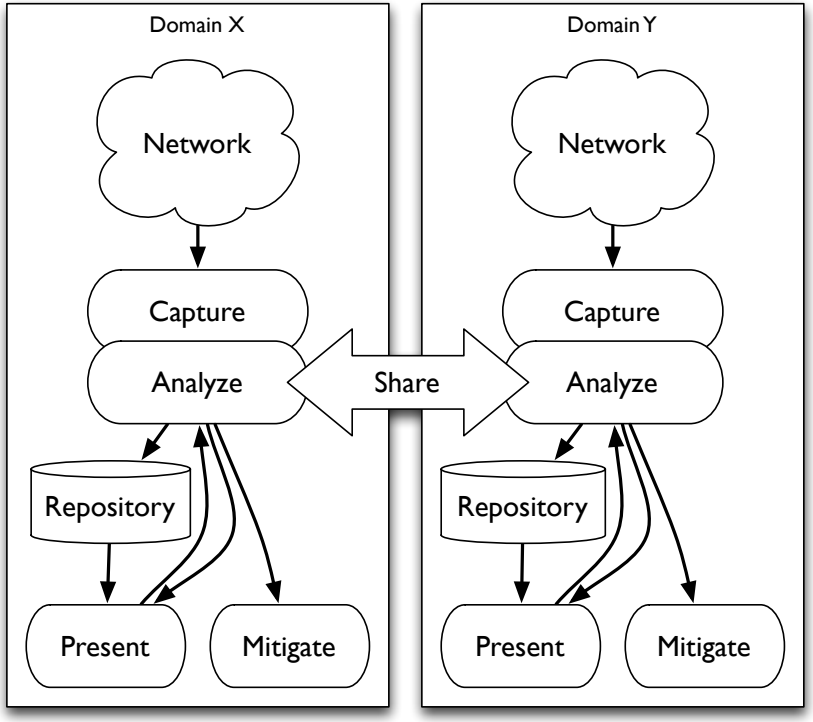
The DEMONS approach: decentralization

- *Move processing to the edge*
- Support iterative analysis on live traffic using programmable edge devices.
- Emphasize stream processing
  - Assumes temporal non-uniqueness of interesting activity in darkspace.
- Data reduction improves scalability.
  - (and reduces sensitivity of collected data)



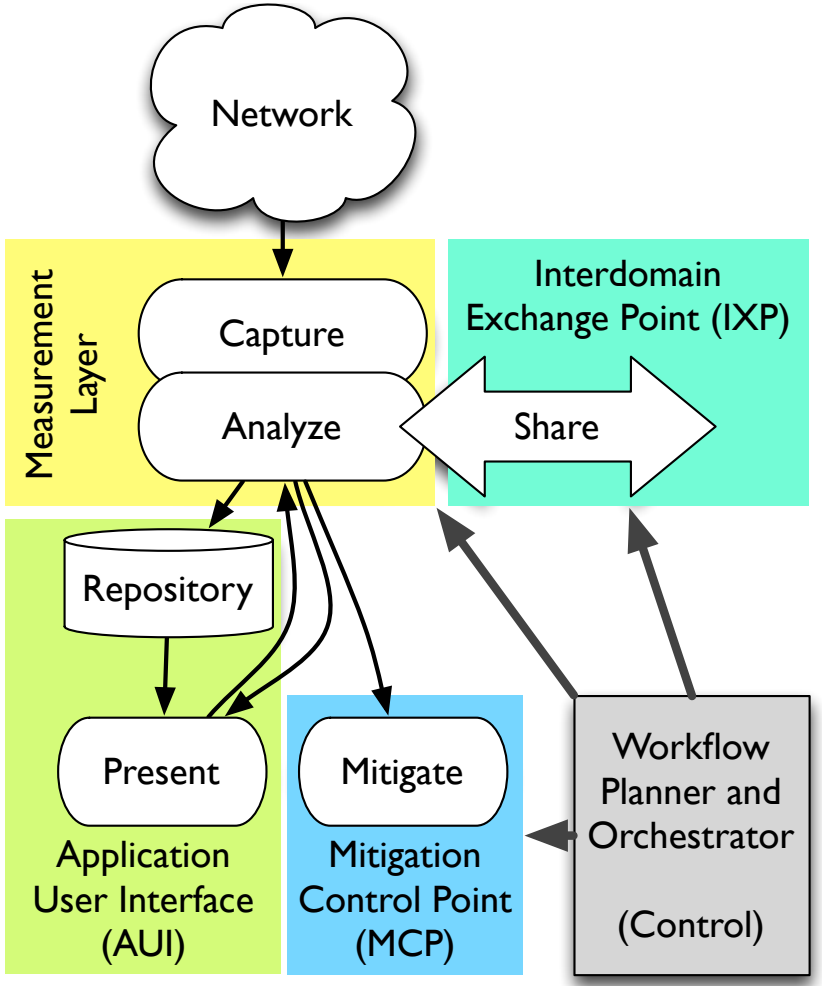
The DEMONS approach: sharing

- *Share analysis, not data*
- Analyses built by composition of well-defined processing modules
- Inspection of intermediate results before export
  - "yes/no decision" instead of anonymization/protection at the edge.
- Realism about technical limitations of data protection
  - Designed to operate within a consortium governed by a legal agreement.



Components and Interfaces

- Measurement layer nodes provide capture and analysis.
- Interdomain exchange point (IXP) provides "sharing" interfaces to external domains.
- Workflow-based control interface, for passing requests within and among domains
- Integration with existing mitigation processes
  - not particularly relevant for research



Inspirations

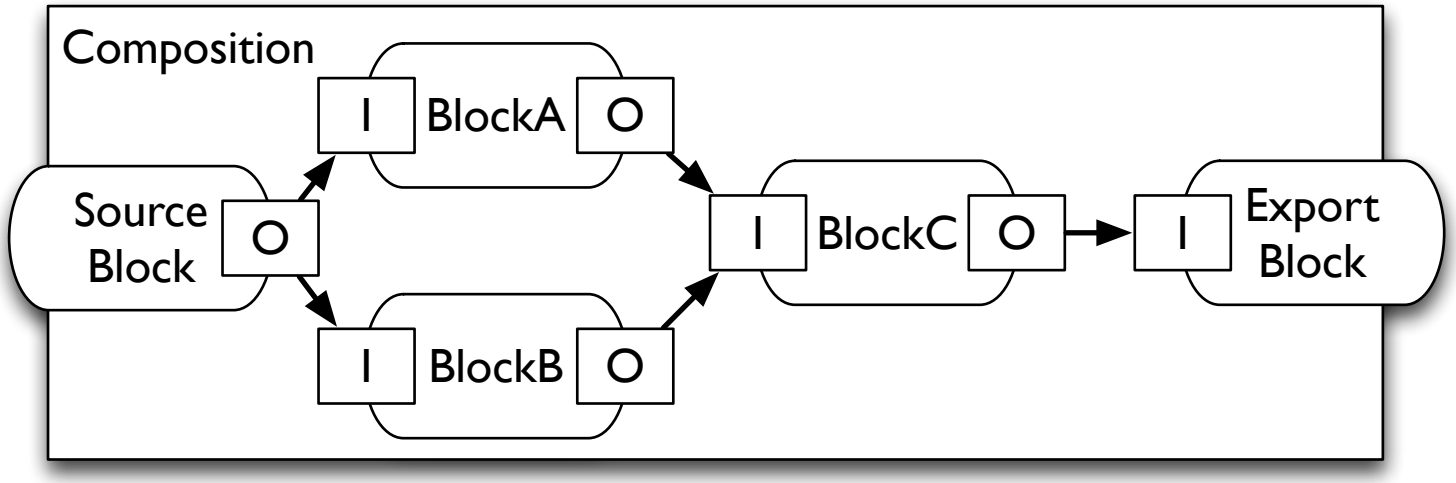
- EU FP7 PRISM (March 2008 – May 2010)
  - Architecture for privacy-aware network monitoring applying context-sensitive access control
  - Explicit application of regulatory framework to access control and conditional data protection
  - Focus on single domain (no sharing)
- Secure Queries (Mirkovic, NDA '08)
  - Trol (SQL-like) language for network queries
  - Anonymity-aware access control
- Dialing Privacy and Utility (Kenneally and Claffy, IEEE S&P July '10)
  - Combines technical data protection techniques with a policy framework considering legal, social, privacy and utility aspects of sharing.
  - Applied to network telescope operations

Measurement Layer: Blockmon

- Composable measurement using small blocks
  - Increases parallelizability, measurement performance on multicore hardware
  - Code reuse for measurement development
- Platform for understanding composable measurement application development
- Enable code and analysis interchange in the form of compositions of modules from a standard, trusted base.
- Current work in ontology of basic operations:
  - Filters, Metrics, Features, Correlations, and Feedback
- Current work in core performance tuning:
  - cache-aware allocation, automatic queue balancing

Blockmon: Implementation

- Compositions of *blocks* exchange *messages* (packets, flows, etc.) connected at runtime via *gates.*
- Control over thread/CPU assignment and block invocation for tuning.
- Blocks, framework and scheduling implemented in C++
- Python-based CLI and JSON-RPC daemon, compositions in XML
- Compositions split among instances via IPFIX import/export



Blockmon: Applicability

- Researchers and data providers agree on a set of blocks as primitives from which to build analyses.
- Additional blocks may be distributed as object code signed by a trusted third party and dynamically loaded at runtime.
- A given study is handed to data provider in terms of a composition.
- Composition runs over live traffic from telescope, or replayed traces / pre-analyzed data kept by the data provider.

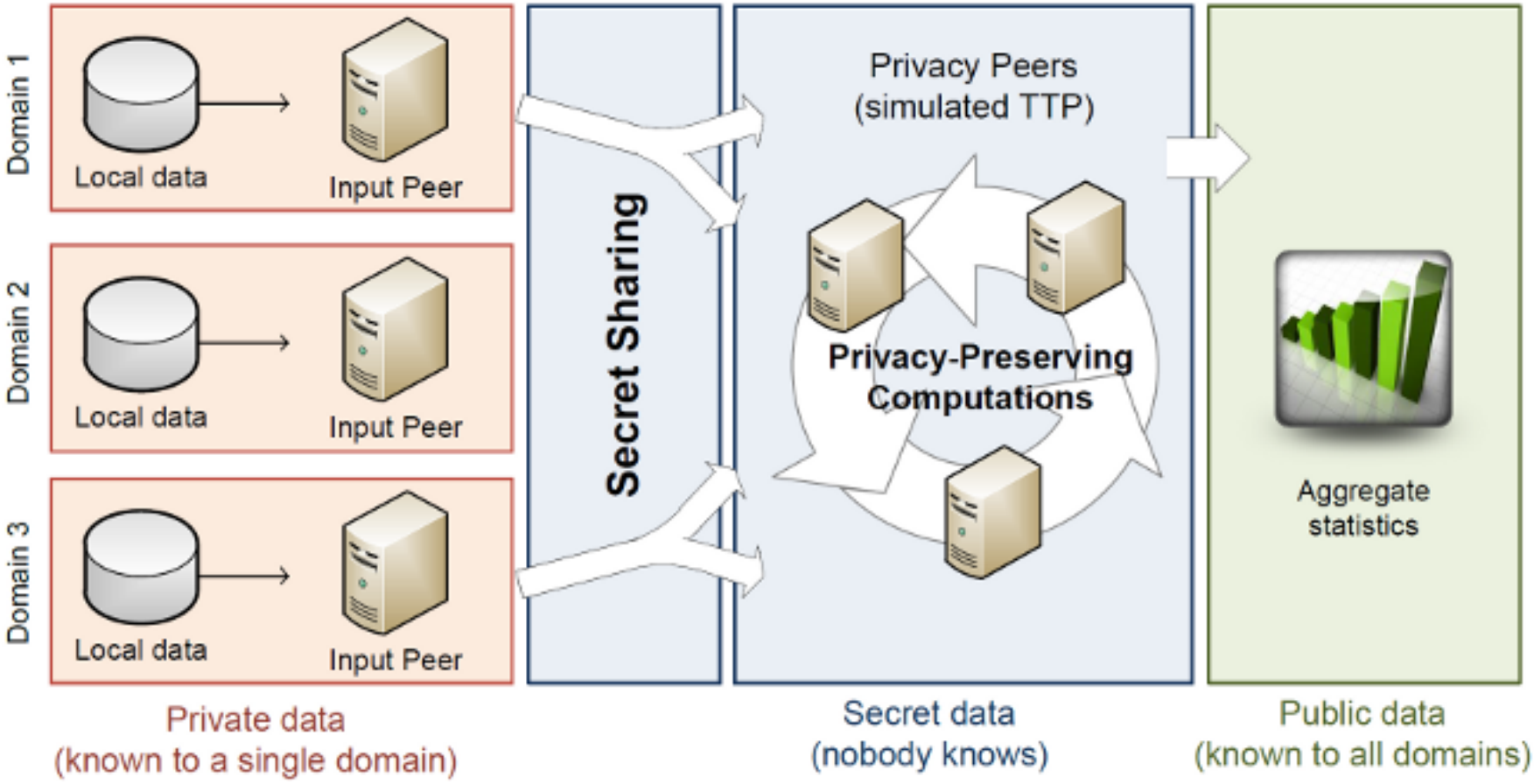
Interdomain Data Exchange

- Intermediate and final results of analyses sent to remote domains through an interdomain exchange point (IXP)
  - Data crossing domain boundaries always significantly reduced for scalability as well as privacy reasons.
- Data subject to inspection at a proxy.
  - Forwarding logic is presently application-specific.
- Data may be further analyzed by another Blockmon instance in the remote domain.
- In certain circumstances, data protected by secure multiparty computation protocols such as SEPIA instead.

Secure Multiparty Computation: SEPIA

- Secure multiparty computation framework based on Shamir secret sharing.
- Optimized for network traffic analysis and list/set operations.
  - can be applied to aggregation over realistically large time bins from reasonably sized networks
  - parallelization of rounds reduces overhead among privacy peers
- Applicable to aggregate or set operations on data from multiple sources without a trusted third party.
- available under LGPL at http://sepia.ee.ethz.ch

Secure Multiparty Computation: SEPIA



SEPIA: Applicability

- Example: aggregate traffic per port without exposing differences per provider among providers.
- Example: union/intersection of unsolicited senders without attribution.
- Limited to situations with
  - many data providers
  - relatively limited data volumes
  - easily defined aggregation operations
- So: unclear whether it's worth the trouble in most network telescope operations.

Guidance for Darkspace Monitoring

- Data sharing is fraught with peril…
  - and becomes moreso the closer you get to lit networks
- …so share analysis instead of data.
  - Composable measurement is one way to do this
  - Sandboxing, code inspection/signing are other approaches
  - A common vocabulary for description (and tools supporting it) would be a third.
- Reduce data aggressively close to the edge
  - Throw away what you *know* isn't interesting.
- Consider "live" analysis
  - Assume that unsolicited traffic is temporally non-unique.

Acknowledgments

- FP7-DEMONS project, especially…
- Blockmon development team
  - NEC Laboratories Europe
  - University of Pisa
  - University of Rome Tor Vergata
  - ETH Zürich