WG Measurement Requirements

Lothar Braun, Thorsten Braun, Georg Carle, Falko Dressler, Anja Feldmann, Dirk Haage (note taker), Tobias Limmer, Tanja Zseby (chair)

Definition

The objective of this working group was to derive measurement requirements and challenges that originate from intrusion detection.

Challenges

First of all, we should be in much better position than other disciplines (like astrophysics, biology, etc.). Our subject of research (the Internet) is accessible. So it is much easier to deploy measurement points in the Internet than for instance in the galaxy. In addition to this, the system is man-made. So it should comprise much more deterministic and predictable behavior than processes in nature. Nevertheless, the participants agreed that the system has already grown too complex comprising to many interconnected processes. In addition, it includes random factors like user behavior and physical influences.

The participants agreed that the ideal measurement system would be one that captures everything everywhere. With this attack detection algorithms (e.g. machine learning) could extract the information needed to detect anomalies. Nevertheless, measuring everything everywhere would mean to fully capture each packet at each network node. This was identified as not feasible by the participants, especially if we don't want to slow down the Internet or make its usage much more expensive.

Since packets travel over multiple nodes, we would get duplicates of each packet for each node. Furthermore, if we also want to secure the network for the collection of measurement results in the same way we consequently would also need to measure the measurement traffic and the measurement of the measurement traffic and so on. This would end up in infinite traffic amplification. Furthermore, we doubt that current IDS systems could handle the high amount of data that would result from full measurements at all nodes.

The main constraints for a measurement system nowadays are:

A) Resource Limitations
   - o Since measurements are only supporting network operation, we are limited in resource that we can afford for network measurements. Limitations apply for processing power, memory and transmission capacity. The situation gets worse in environments with wireless transmission and small mobile devices.

Dagstuhl Seminar Proceedings 08102
Perspectives Workshop: Network Attack Detection and Defense
http://drops.dagstuhl.de/opus/volltexte/2008/1496

B) Privacy
   - Users have a high interest that their privacy is protected. Capturing flow data and especially packet content (header, payload) clearly violates this. Anonymization techniques have the disadvantage of removing information, which might contradict analysis rules. Providers have an interest that others do not gain knowledge about their network and users. As a consequence getting network traces and sharing of data is extremely difficult nowadays.

Derived from these constraints we define the following measurement challenges:

The main challenge of course is to predict the future. For new attacks we only know afterwards what data would have been useful to detect them. In addition to this, we have to apply measurements in different environments with different constraints.

In order to cope with resource constraints we need to look at

- Resource Management
  - Provide more resources: Provide faster hardware
  - Reduce Resource Requirements
    - Provide improved algorithms for measurement tasks (e.g. for capturing, classification)
    - Apply smart aggregation and data selection techniques
  - Resource Adaptation
    - Provide flexible and controllable measurements to control resource consumption
    - Avoid undesired behavior in the control loop (oscillations)
- Data Sharing
  - Provide reference trace for research community, in ideal case labeled data
  - Incentives to network operators to share data
  - Standardized measurement methods and result representation to ensure comparability of measurement results
  - Privacy-preserving methods
  - Methods to bring analysis code to the data instead of data to the code
- Data processing
  - association of data from different layer and observation points, e.g. Information from the end host or meta information
- Protection of the system (against attacks, compromises, overload)
- Dealing with encrypted traffic

From this also challenges can be derived for Intrusion detection system:

- Work with incomplete data
- Specify in detail which data is required
- Develop methods to take advantage of adaptable measurement

Assessment of state of the art

There are a wide variety of measurement tools for packet and flow measurement available. The MOME database (www.ist-mome.de) lists more than 400 different tools. Configurable measurements are possible to a certain extend. Standardization efforts are in progress (IETF IPPM, IPFIX, etc.).

We are not able to provide labeled data. Data from existing networks is hard to classify in attack/no-attack traffic because we don't know the truth, i.e., we do not know exactly whether the traffic may include new attacks or not. Due to the complexity of nowadays networks, generating artificial traces is extremely difficult. But data from honeypots or honeynets may help to collect useful input for training IDS systems.

Future directions and Recommendations

First of all, intrusion detection and measurement community should work more closely together. We provide tools and know how to improve and to specialize them. It is useful to get requirements for this from the users of the data. Specific questions to users of data (e.g. machine learning algorithms, etc.) would be the following:

- What data do you need? (if you cannot get all)
    - Cover multiple observation points/multiple flows with coarse grained measurements OR details for a few flows/packets/observation points
    - Can you work with incomplete data sets, sampled, aggregated,…?
    - How relevant is the correlation of data?
- Can re-configurability help?
    - If you see some data do you know what you need next? or what you don't need…
    - How often do you need to reconfigure?
- What are time requirements?
    - How frequent do you need updates?
    - With how frequent updates can you cope?

It is important to gather cross-layer and meta-information and associate it with pure measurement results. This includes information about specific events that may influence traffic, measurements from end system and multiple layers.

In addition, the participants agreed that measurement should be an integral part of future network design and that we would prefer a single measurement system for different task (intrusion detection, accounting, SLA validation). Since we currently do not yet see that one of the current approaches takes off we cannot yet design the measurement infrastructure and still need to observe the developments in this community.

4 **Lothar Braun, Thorsten Braun, Georg Carle, Falko Dressler,** Anja Feldmann, Dirk Haage (note taker), Tobias Limmer, Tanja Zseby (chair)