# Design and Implementation of an Anomaly Detection System

Luca Deri <deri@ntop.org>
Gaia Maselli <maselli@di.unipi.it>
Stefano Suin <stefano@unipi.it>

# Goal of This Work

- In every network there are some global variables that can be profitably used for detecting network anomalies, regardless of the type of network users and equipment.

- As most of the relations among these variables are fixed, it is possible to define generic network rules for automatically detecting selected network anomalies.

# How N-IDS Systems Work [1/2]

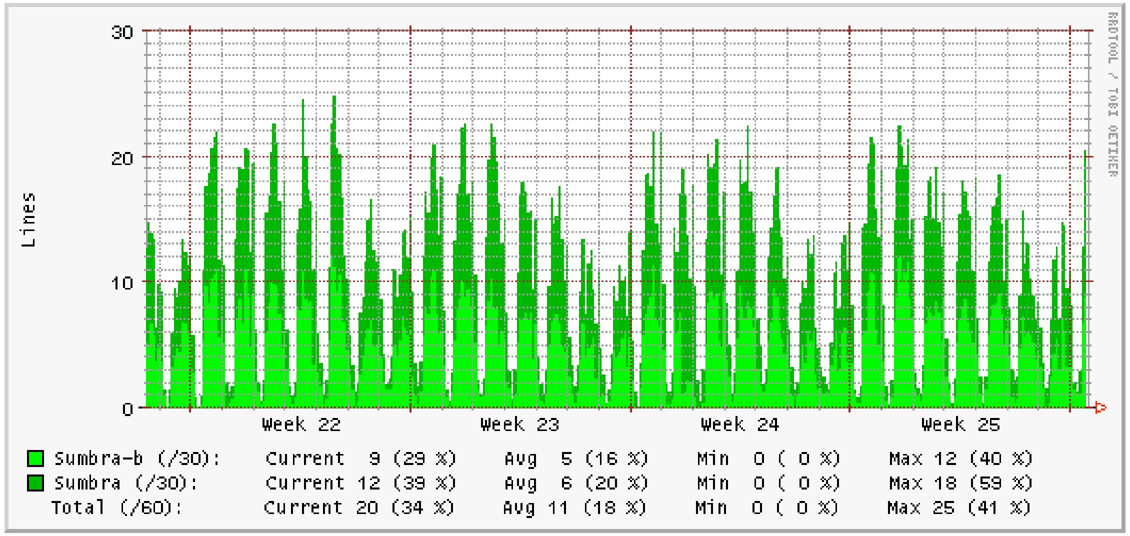
- Signature detection systems use patterns of well-known attacks or weak spots of the system to match and identify known intrusions.
- Advantage: known attacks are detected efficiently.
- Disadvantage: lack of the ability to detect new attacks

# How N-IDS Systems Work [2/2]

- Anomaly detection systems flag observed activities that deviate significantly from the established normal usage profiles as anomalies: something that is abnormal is probably suspicious.
- Advantage: it does not require prior knowledge of the intrusion so it can detect new intrusions.
- Disadvantage: no clear definition of an attacks hence it can have high false positive rate.

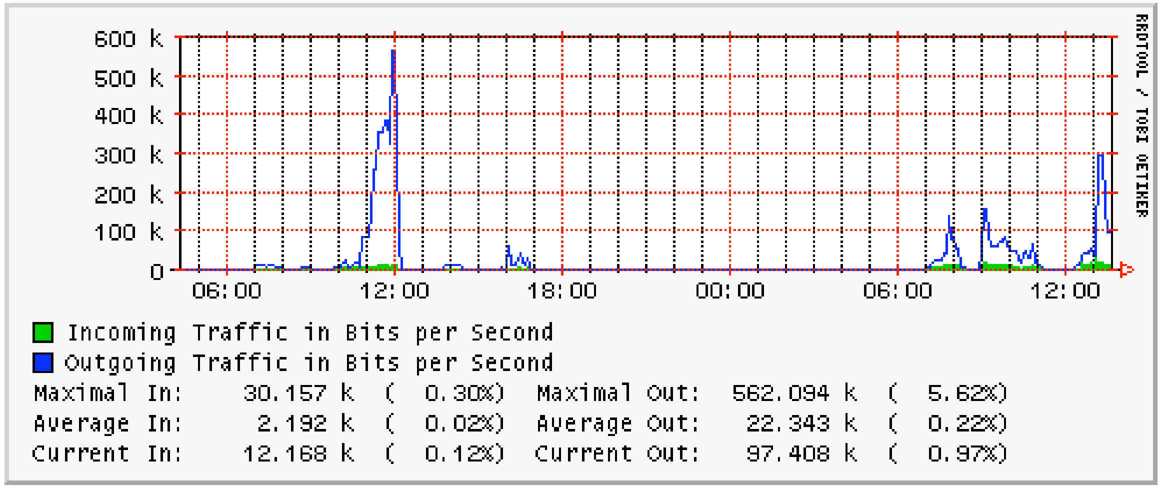
# Defining a new Type of Anomaly Detection System [1/3]

- Various experiments performed on different networks confirmed the presence of some similarities on traffic.



# Defining a new Type of Anomaly Detection System [2/3]

- Simple bytes/packets curves are not very reliable for detecting networks problems, as they can present some peaks caused by various reasons (e.g. a multicast transmission).



🟩 Incoming Traffic in Bits per Second
🟦 Outgoing Traffic in Bits per Second

```
Maximal In:     30.157 k  (  0.30%)  Maximal Out:   562.094 k  (  5.62%)
Average In:      2.192 k  (  0.02%)  Average Out:    22.343 k  (  0.22%)
Current In:     12.168 k  (  0.12%)  Current Out:    97.408 k  (  0.97%)
```

# Defining a new Type of Anomaly Detection System [3/3]

The authors decided to investigate whether it was possible to:

- Identify some selected traffic parameters that can be profitably used to model network traffic behaviour.
- Define traffic rules so that when such rules are violated there is necessarily a network anomaly (e.g. an abnormal network activity).

# What is an Anomaly?

The deviation from the network's expected behaviour that is defined by considering two kinds of knowledge:

- IP protocol specifications contained in RFCs, that needs to be satisfied by every host and network (static knowledge).
- Statistical traffic analysis that varies according to network characteristics and type of users (dynamic knowledge).

# Building Static Knowledge

- Classification of effects on the network of known network security violations.
- IP protocol dissection (RFCs).
- Network traffic monitoring parameters used by monitoring applications (e.g. RMON)
- Experience: survery of parameters checked by network administrators

# Building Dynamic Knowledge

Produce a traffic model for each monitored asset that includes:

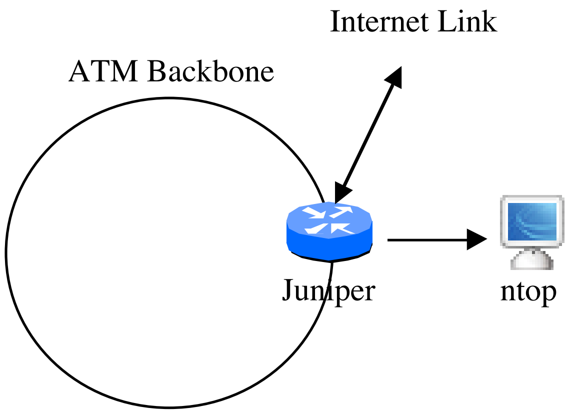
- List of provided network services.
- Thresholds for some specific traffic (e.g. SYN pkt ratio, # concurrent outgoing connections).
- A security index that idenfies how "safe" is an host.

# Some Common Traffic Parameters

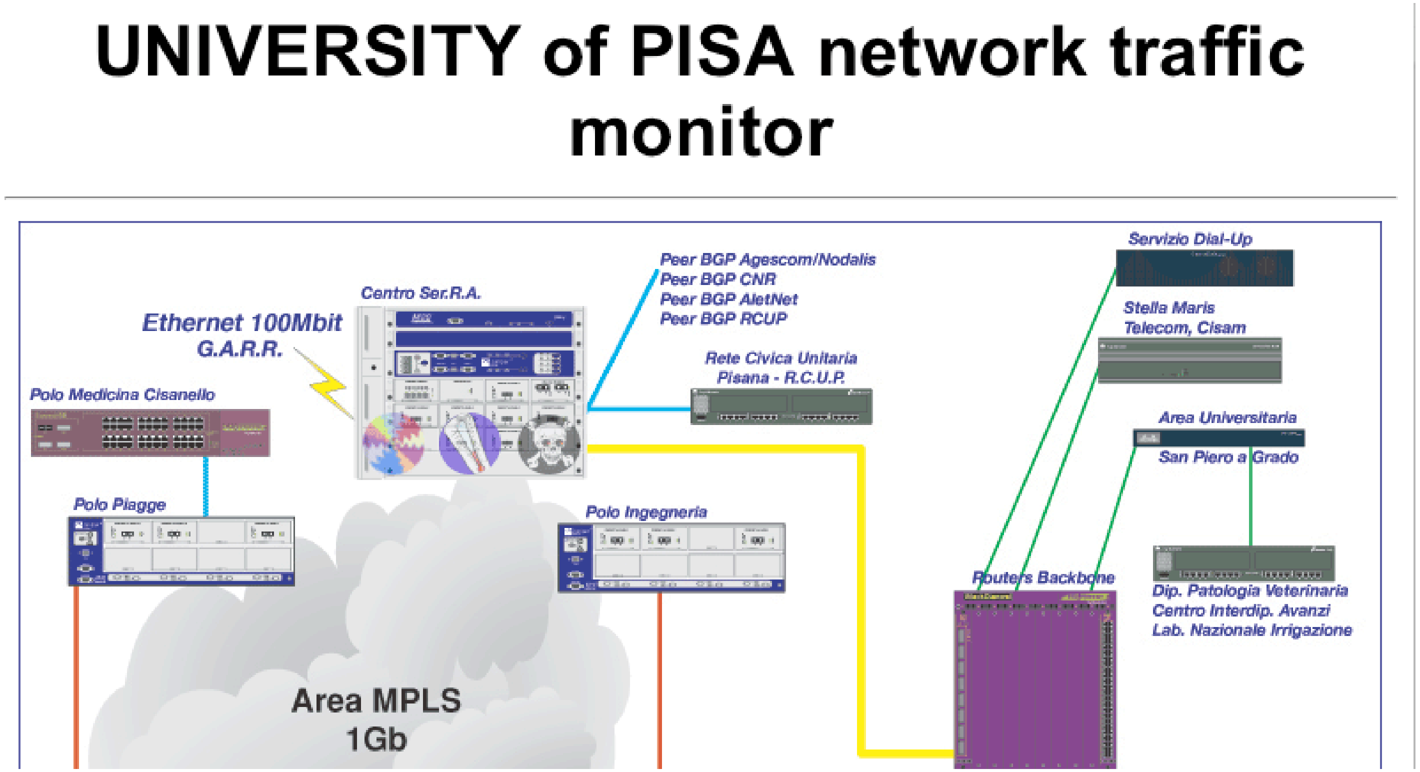
- ICMP ECHO request/response ratio
- ICMP Destination/Port Unreachable
- # SYN Pkts vs. # Active TCP Connections
- Suspicious Pkts (e.g. out of sequence)
- Fragments percentage
- Traffic from/to diagnostic ports (e.g. ident)
- TCP connections with no data exchanged

# Validation Playground [1/2]

- Extension to ntop for accounting selected traffic parameters and calculating security thresholds.
- Test on the Unipi backbone



# Validation Playground [2/2]



UNIVERSITY of PISA network traffic monitor

http://mrtg.unipi.it

# Evaluation [1/3]

- Anomaly detection based on expected behaviour and the study of RFCs, guarantees a better longevity with respect to detection mechanisms based on pattern matching and signature detection.
- The ADS is effective in many situations where a firewall or an intrusion detection system fail (e.g. a cracker gain host access by means of a buffer overflow).
- Attacks, when classified in terms of anomaly categories, are very few with respect to the large number of signatures and patterns that similar solutions need to handle.

# Evaluation [2/3]

- # of knowledge rules you use:
  - ~50 rules per host
  - ~20 global rules (applied to the whole net)
- Rate of false positives
  - < 10% on "known" hosts
  - "unknown hosts": investigation needed (informational)
- What is normal behaviour?
  - Thresholds for servers/workstations/p2p's

# Evaluation [3/3]

The study of the results produced by the ADS can be very well used for:

- Network bandwidth optimisation.
- Detection of network bandwidth killers.
- Avoidance of unwanted protocols.
- Network misconfiguration.
- Unwanted server activity detection.
- TCP/IP stack tuning based on the distribution of TCP connection number, flags (e.g. RST, SYN), and latency.

# Ongoing Work

- pTop (S.Suin and D.Vaghetti):
  - Realtime traffic analyzer that is activated on demand when a potential problem is detected
  - Data storage on a MySQL database
  - Web interface for access to monitoring data