Packet Sampling and Network Monitoring

CERN openlab Monthly Technical Meeting

13th November, 2007

Milosz Marian Hulboj

milosz.marian.hulboj@cern.ch

Ryszard Erazm Jurga

ryszard.jurga@cern.ch

What is “Network Monitoring”?

- Network “Health” Inspection
- Observation and analysis of following objects:
  - Network devices
  - End systems
  - Network links
  - Network traffic
  - Network applications

Why Network Monitoring (1)

- Networks are getting more complex and harder to comprehend
- Networks are a business-critical element
- Occurrence of problems in any network is inevitable:
  - Increasing configuration and topology complexity
  - Increasing number and complexity of threats, attacks, viruses, etc.
  - Conclusion: It is just a matter of time
- Detect the problems as early as possible
- Reduce the unavailability time

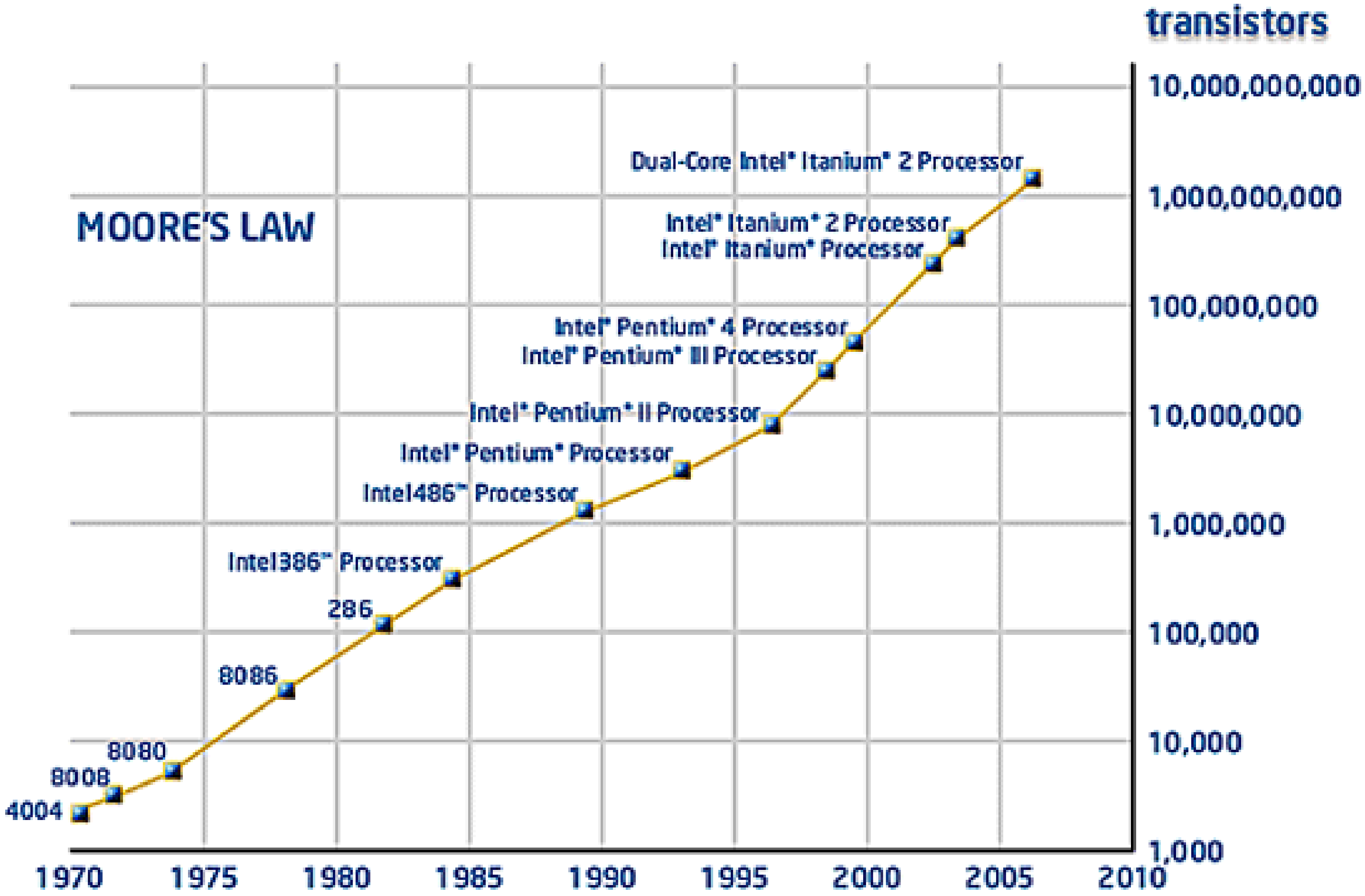
Why Network Monitoring (2)

- Network Statistics:
  - Identification of performance characteristics:
    - For traffic engineering (pkt/s, bytes/s, connections/s, flows, traffic matrix)
    - QoS metrics, latency, bandwidth (SLA, billing)
    - Planning (busiest services, traffic distribution, throughput)
- Network Inventory:
  - Identification of equipment on the network
- Troubleshooting:
  - Failures of interface cards, power supplies
  - Connectivity problems
  - Service availability

Why Network Monitoring (3)

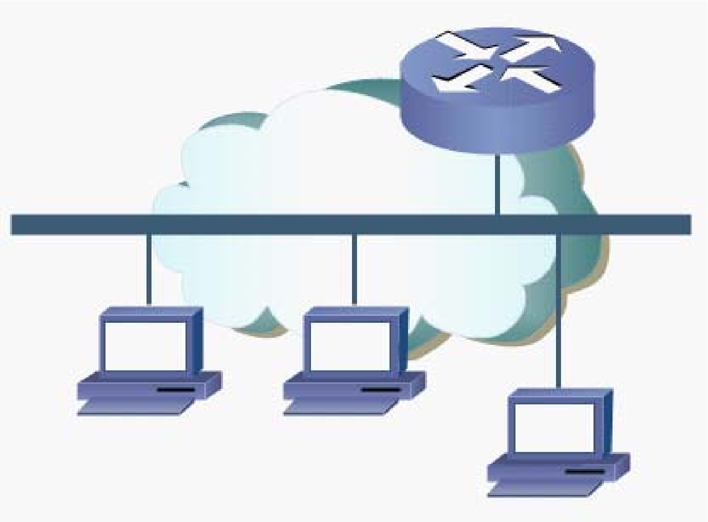
- Accounting
  - User activity
- Security
  - Policy violations:
    - Unauthorised services, machines
    - Unauthorised access
    - Unauthorised applications (e.g. p2p)
  - Intrusion detection
  - Compromised hosts detection
  - Protection against cyberattacks, worms, etc.

Performance Problem

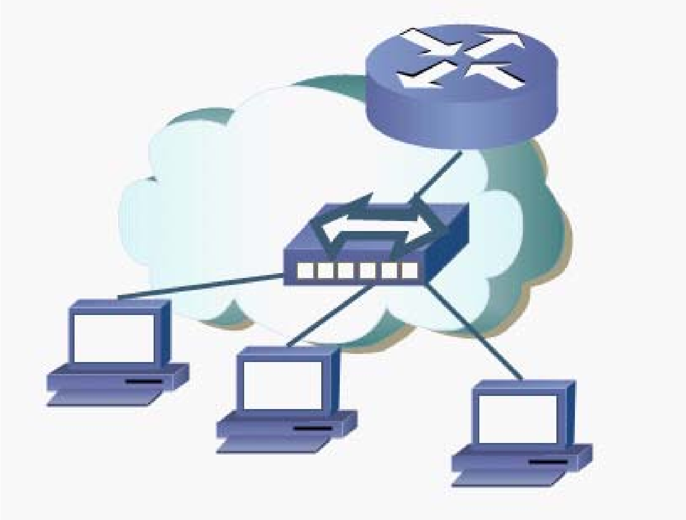


Packet Analysis - old methods (1)

- Sniffing in the old times (“old shared Ethernet”)
- Slow network speed
- Captures everything (all packets+payload)

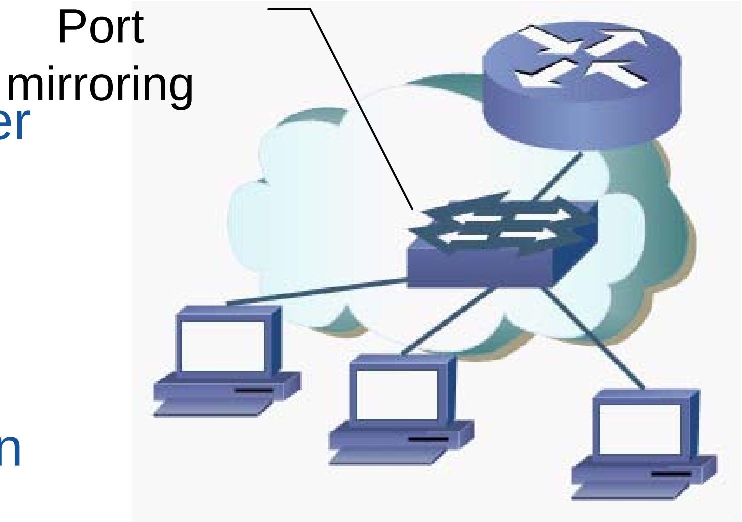


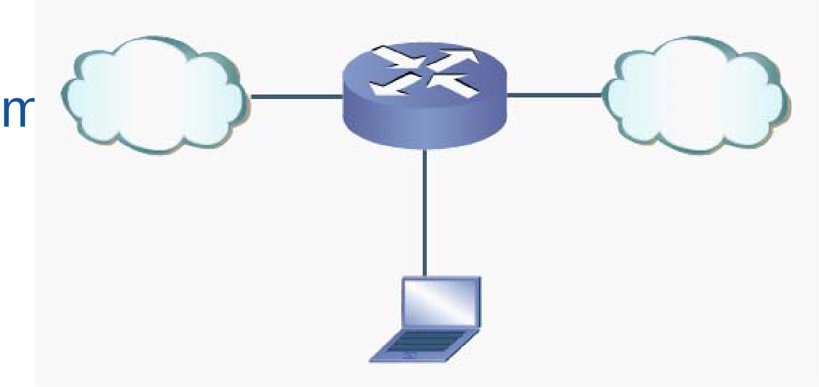
- “Old shared Ethernet” is a history…



Packet Analysis – modern methods (2)

- Port mirroring:
  - Captures all the traffic (per port, group, VLAN, etc)
  - Requires HW support
  - Requires fast network interface
  - Problematic determination of originating port
- Network device-based data:
  - Captures (partial) data from selected ports
  - Sampled packet data
  - Sampled flow data
  - Requires HW support





Other Common Sources of Data (1)

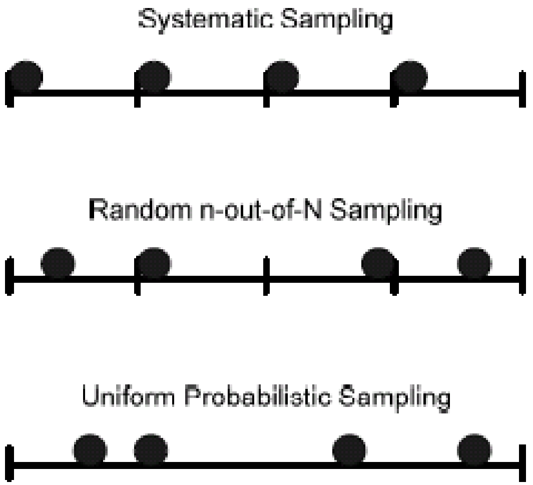
- SNMP
  - Operations via simple variable manipulation
  - Standard mean for retrieving generic statistics, network status, etc:
    - Packet arrival and departure rates, packet top rates, error rates, system load, etc.
  - Used also for network configuration
  - Cannot customise monitored variables within agent
  - Different vendors use different proprietary MIBs for detailed information

Other Common Sources of Data (2)

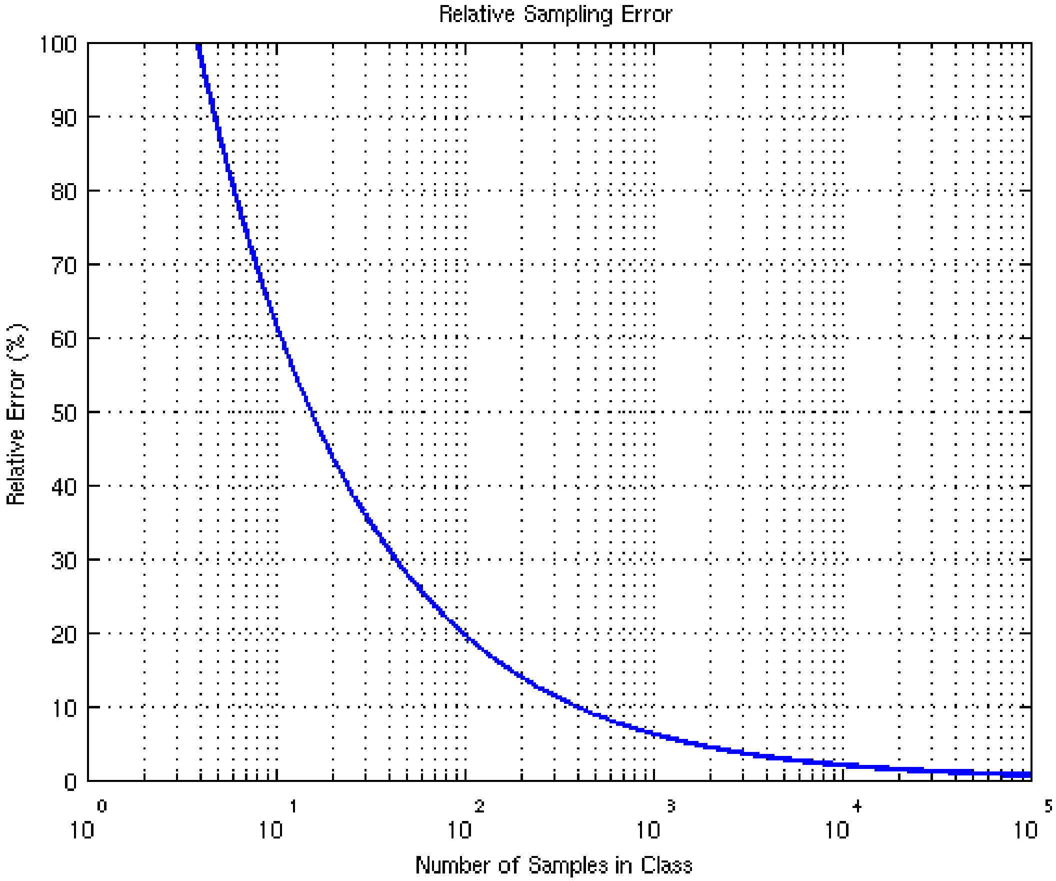
- RMON and RMON2
  - Extension of the basic set of SNMP
  - Remote data collection and processing
  - RMON2 decodes packets at layers 3 – 7 and handles certain protocols
  - Collects aggregate statistics (volume, rate, Top Talkers, etc) about network and application traffic
  - Implementation of RMON agents is complex
  - Probes might be expensive and require administration
  - Cannot add new features to the existing MIB

Packet Sampling

- A mean of passive network monitoring
- Simple to implement
- Low CPU and memory overhead
- Sample only the packet header (~128 bytes)
- Traffic patterns estimated from the samples with certain error



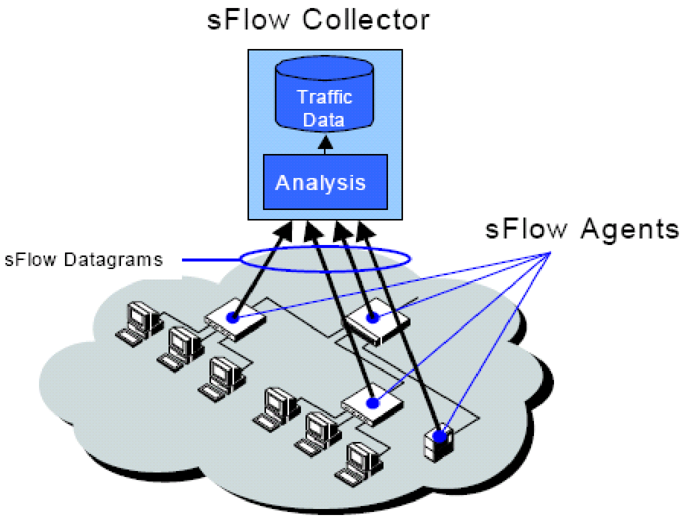
Error Estimation



- Decreasing error = increasing the sampling rate

sFlow Packet Sampling

- RFC 3176
- Multi-vendor standard
- Complete packet header and switching/routing information
- Some SNMP counters information
- Low CPU/memory requirements – scalable



Usage of sFlow

- Profiling network traffic
- Building flow statistics
- Accounting and billing
- Route profiling (forwarding information)

- Security analysis / intrusion detection:
  - Packet headers analysis
  - Traffic pattern analysis

Current Research

- Influence of sampling on flow estimates
- Influence of sampling on anomaly detection:
  - Access only to packet headers
  - Unable to reconstruct the sessions from samples
- Traffic prediction:
  - Packet count prediction
  - Traffic volume prediction
- Adjusting of sampling rate:
  - Attempt to maintain the constant error
  - Attempt to fully utilise the hardware capabilities

Bibliography

- N.Duffield, “Sampling for Passive Internet Measurement: A Review”, Statistical Science 2004, Vol. 19, No. 3
- http://www.sflow.org
- J.Jedwab, P.Phaal, B.Pinna, “Traffic Estimation for the Largest Sources on a Network, Using Packet Sampling with Limited Storage”, HP Laboratories Bristol, HPL-92-35, 1992
- B.Choi, J.Park, Z.Zhang, “Adaptive Random Sampling for Load Change Detection”, University of Minnesota Technical Report 01-041, 2001
- B.Choi, J.Park, Z.Zhang, “Adaptive Random Sampling for Flow Volume Measurement”, University of Minnesota Technical Report 02-040, 2002
- K.Ishibashi, R.Kawahara, M.Tatsuya, T.Kondoh, S.Asano, “Effect of Sampling Rate and Monitoring Granularity on Anomaly Detectability”, IEEE Global Internet Symposium, 2007
- R.Kawahara, T.Mori, N.Kamiyama, S.Harada, S.Asano, “A Study on Detecting Network Anomalies Using Sampled Flow Statistics”, IEEE International Symposium on Applications and the Internet Workshops 2007
- Y.Gao, G.He, J.Hou, “On Exploiting Traffic Predictability in Active Queue Management”, in Proceedings of IEEE INFOCOM, 2002
- N.Duffield, C.Lund, M.Thorup, “Estimating Flow Distributions from Sampled Flow Statistics”, Proceedings the ACM SIGCOMM Conference on Applications, Technologies, Architectures, 2003
- K.Xu, Z.Zhang, S.Bhattacharyya, “Profiling Internet Backbone Traffic: Behavior Models and Applications”, SIGCOMM Comput. Commun. Rev., Vol. 35, No. 4. (October 2005), pp. 169-180.
- And many more…