Network Monitoring
&
Deep Packet Inspection

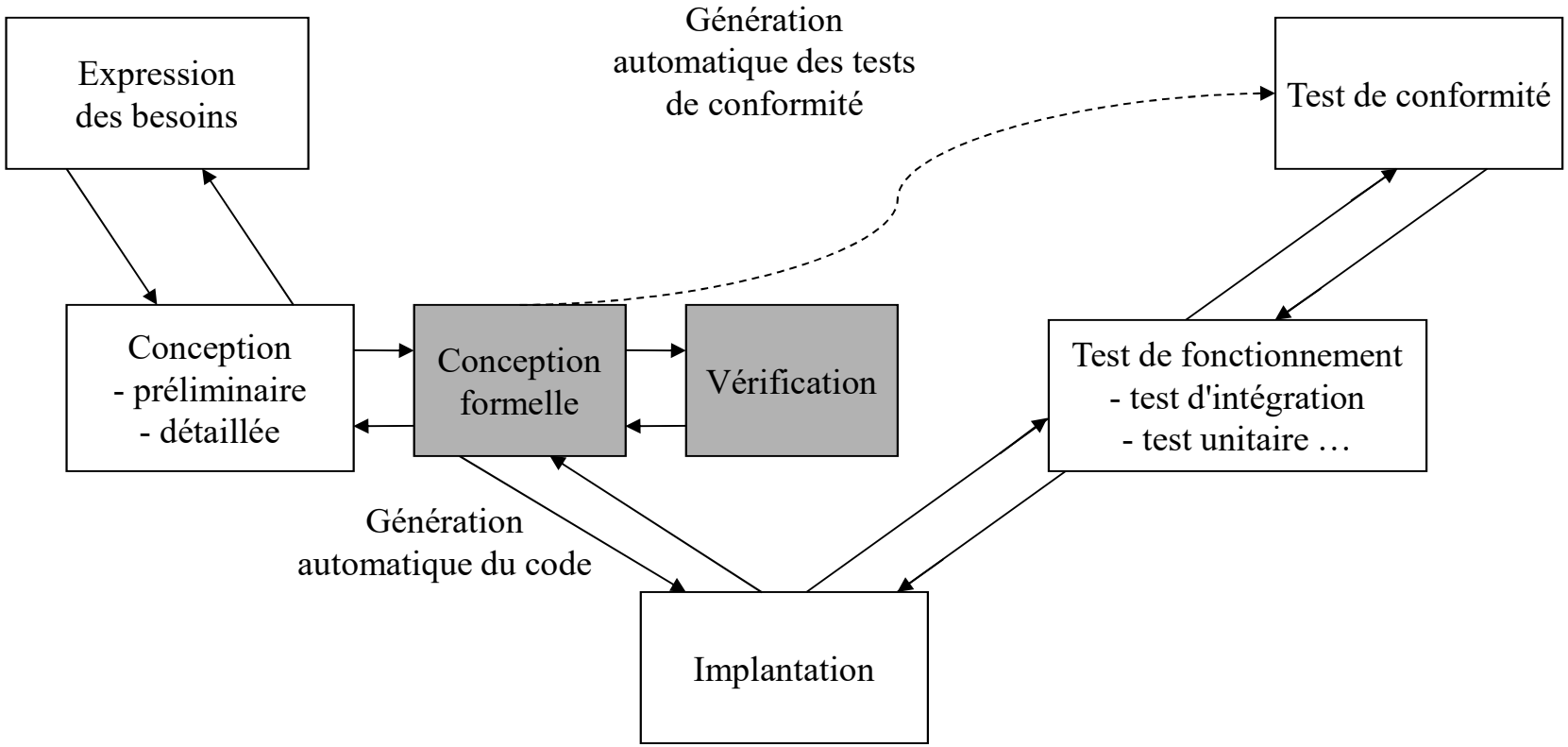
Wissam Mallouli

wissam.mallouli@gmail.com

Personal Background

- Expertise in
  - Formal methods
  - Protocol design & engineering
  - **Monitoring techniques**
  - Evaluation and optimization
- Working on several European research projects
- →Still coding from time to time (between us)

Protocol development lifecycle



Ingénierie des protocoles

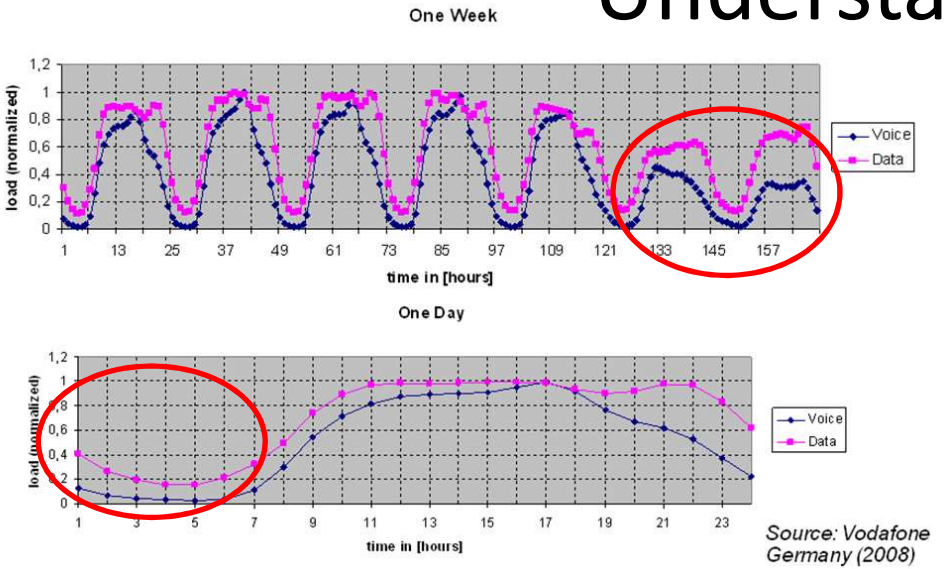
- Les différentes étapes du cycle de vie des protocoles
  - Expression des besoins
  - Conception
  - Description Formelle
  - Vérification
  - Implantation
  - Test et validation
  - Déploiement
  - Maintenance

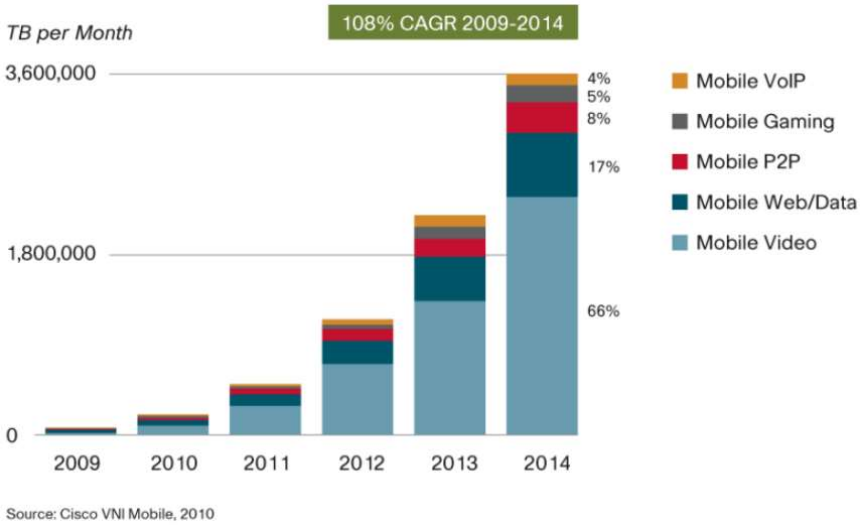
développement

Seminar plan

- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

Need for network monitoring
Understand / Plan



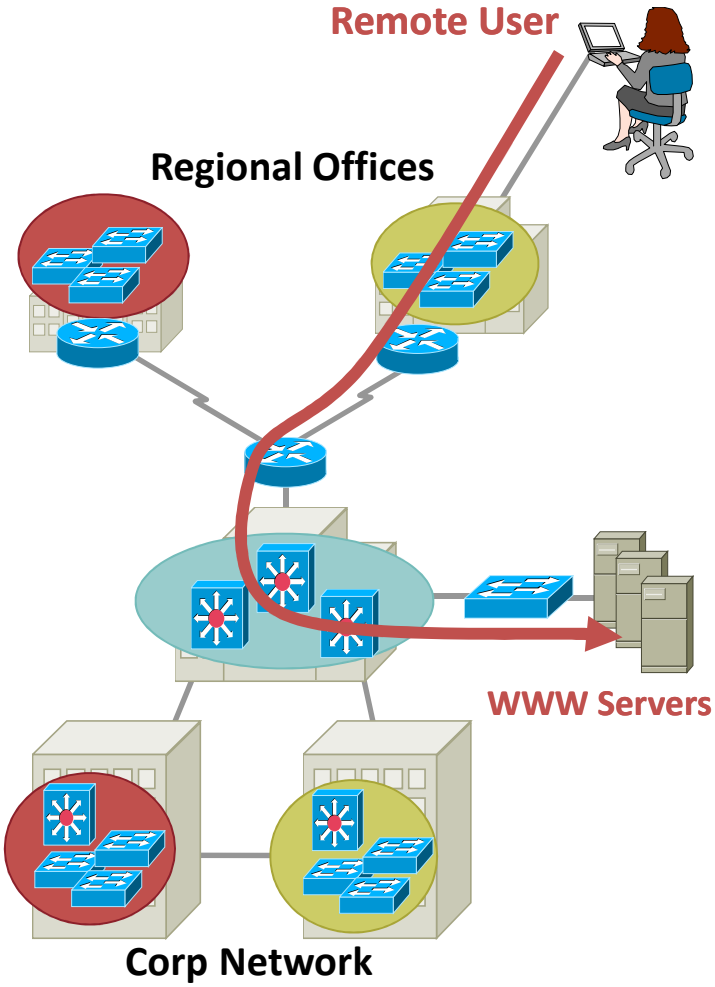


- Need to have a clear visibility over the network
- Status, traffic trends, peak time, evolution, etc.
- Aggressive pricing the week-end/night

- 3.5 billion mobile broadband users by 2015
- Traffic increase 30 times (wrt 2010)
- Understanding
  - the drivers of this growth
  - applications/usages
  - → contribute for a successful network planning

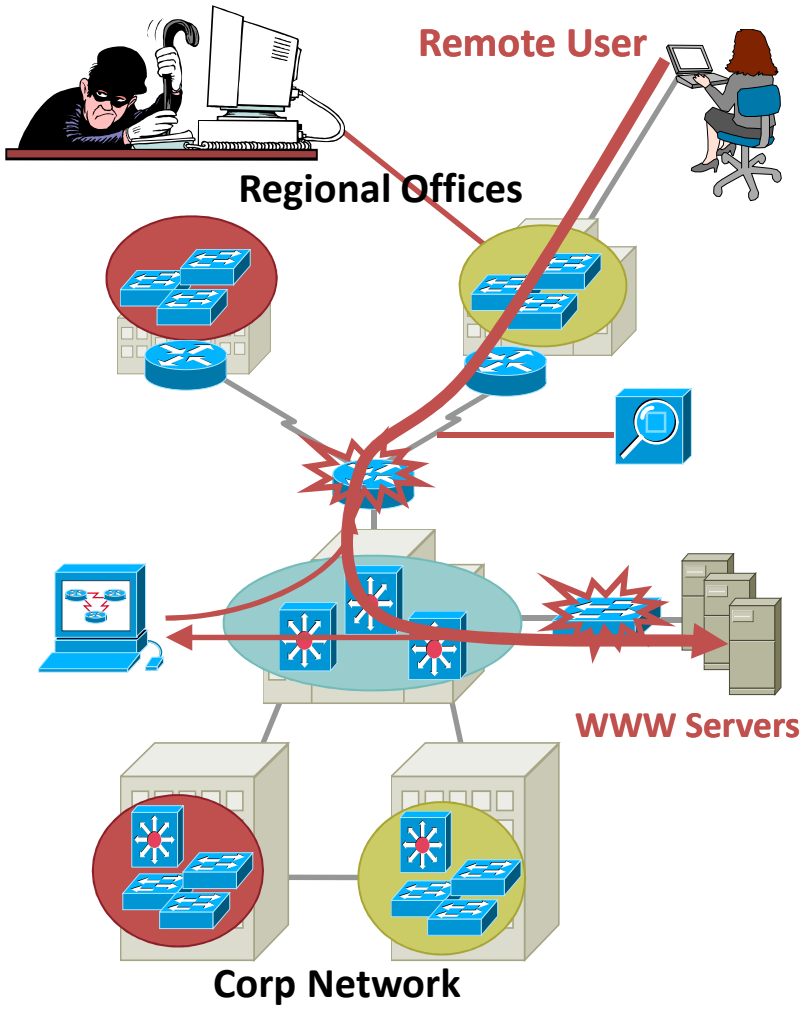
Need for network monitoring
Diagnose & react

- Typical problem
  - Remote user arrives at regional office and experiences slow or no response from corporate web server
- Where to begin?
  - Where is the problem?
  - What is the problem?
  - What is the solution?
- Without proper network monitoring, these questions are difficult to answer



Need for network monitoring
Diagnose & react

- Typical problem
  - Remote user arrives at regional office and experiences slow or no response from corporate web server
- Where to begin?
  - Where is the problem?
  - What is the problem?
  - What is the solution?
- Without proper network monitoring, these questions are difficult to answer



Need for network monitoring
Problems might still occur!

The Amazon Web Services Outage (April 2011)

- Elastic Block Store: storage database for Amazon's EC2.
  - EBS clusters (Database nodes)
  - Control Plane Services (accepts user requests and directs them to appropriate EBS clusters)
  - Inter node communication on high bandwidth network and a lower capacity network as a back-up.
- Manual Error with routine Network Upgrade Procedure (traffic directed to the low capacity network)
  - many nodes were isolated.
  - Etc. (for more info follow the link below)
- Tens of sites and businesses impacted for 3 long days

(Source) http://aws.amazon.com/message/65648/

What is network monitoring

- Process of observing or inspecting the network at different points
- With the objective of
  - Drawing operation baselines
  - Produce reports
  - Notify on abnormal operation
  - Provide input to network management

- Can be used to
  - Understand the behavior of the network
  - Detect faults and abnormal operation
  - Network planning & resource optimization
  - Network security (Intrusion & Attack Detection)
  - Performance, quality & SLA monitoring
  - CRM, Marketing
- Sit above traffic measurements
  - Gather traffic measures and performance indicators
  - Analyze and correlate the measures in order to make a diagnosis

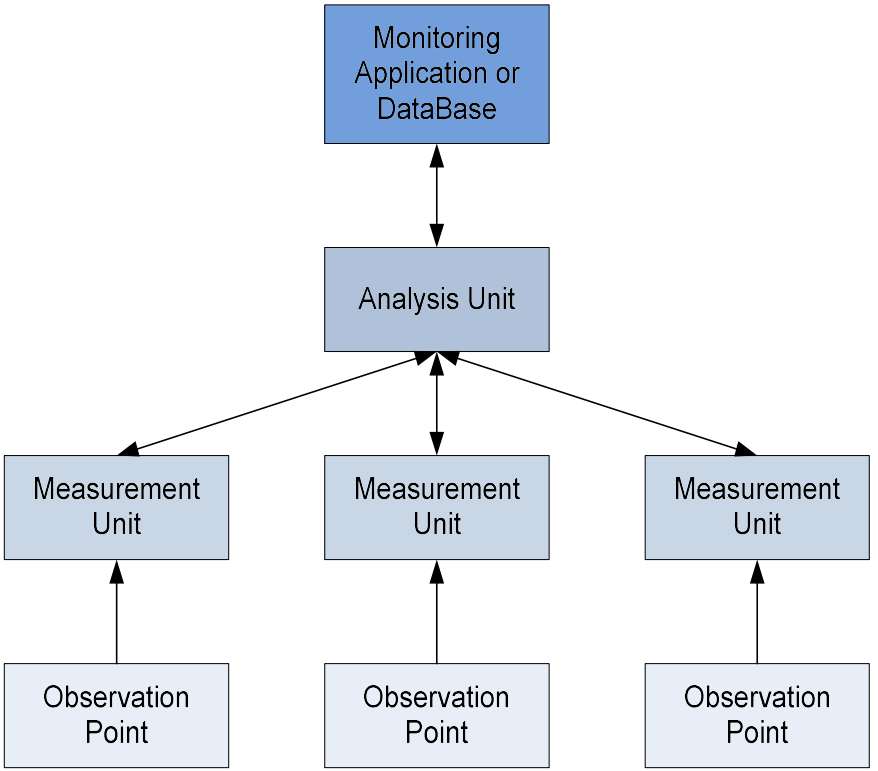
Network monitoring: Basics

Monitoring application
- Interface for real-time monitoring.
- Store collected data in a database for post analysis (trends, history, reporting …).

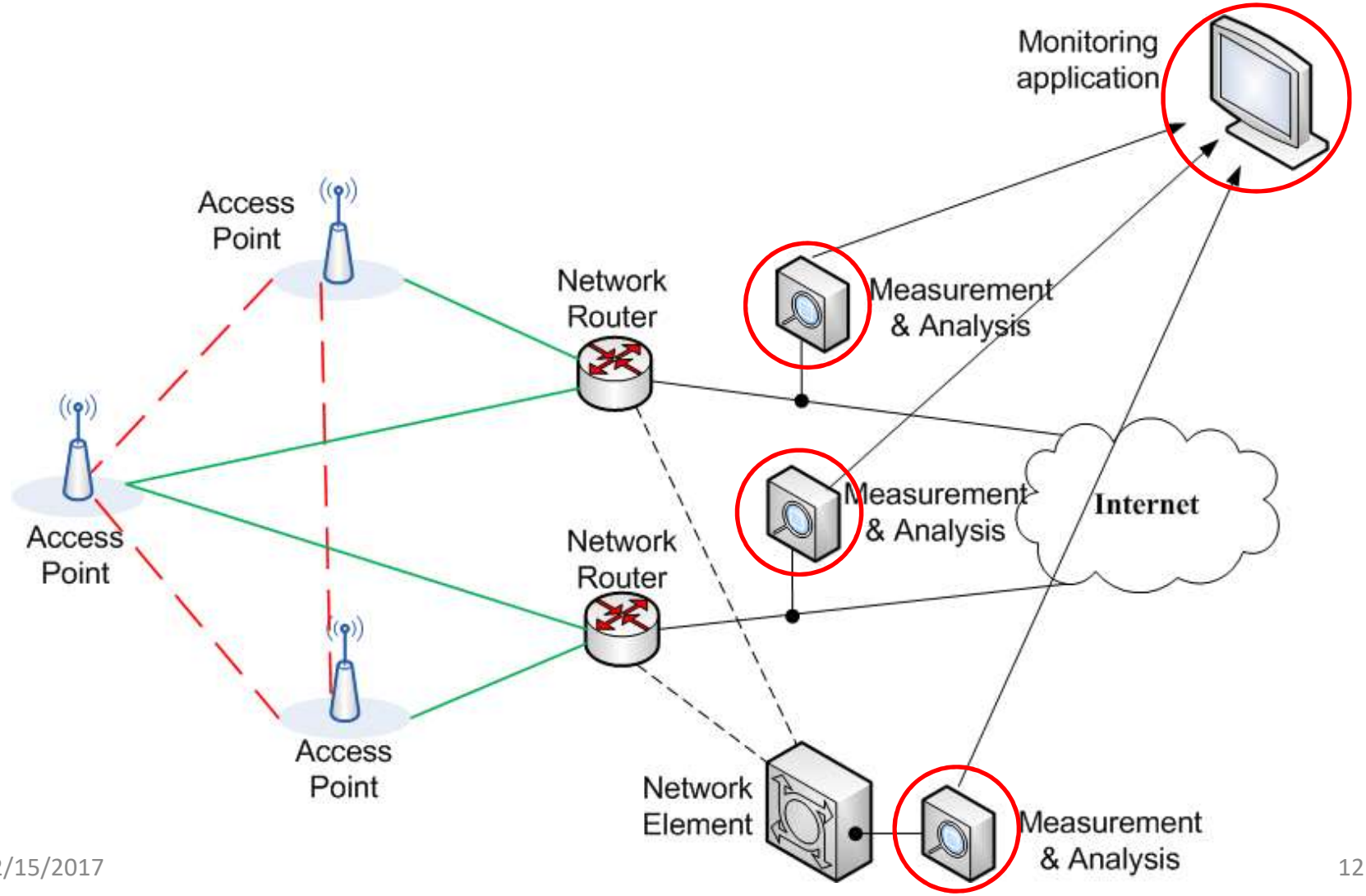
Performance analysis unit
- Gathers, aggregates and correlates collected measures
- Calculate indicators (QoS parameters, KPIs).

Measurement unit
- This performs the measurements
- Collects the required information for analysis.

Observation points
- This is where the measurements will be performed.
- The more observation points we have, the more accurate data we get.



Network monitoring: Basics



Seminar plan

- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

Complexity of network measurements

- Size, complexity and diversity of the networks
  - understand cause-effect relationships is difficult
- Measurement is not an objective!
  - meaningless without careful analysis
  - Analysis depends on the monitoring objective
  - Need to define
    - What, where, how to measure?

Determining *What* to Measure

- Before any measurements can take place one must determine what to measure
- Definition of metrics is closely related to the monitoring objective
- There are many commonly used network performance metrics
  - CAIDA Metrics Working Group (www.caida.org)
  - IETF's IP Performance Metrics (IPPM) Working Group

Determining *What* to Measure

- Example: Performance metrics can be classified into
  - Network metrics
    - Latency
    - Throughput
    - Arrival rate
    - Link utilization, bandwidth
    - Loss rate
  - Application metrics
    - Response time
    - Connection setup time
    - availability
  - User quality metrics (depends on the application)
    - Mean opinion score (VoIP)
    - Quality of experience (Video – through estimation)

Determining *How* to Measure

- Active measurements

- Passive measurements

Determining *How* to Measure

- Active measurements
  - Send test traffic into the network
    - Generate test packets periodically or on-demand
    - Measure performance of test packets or responses

  - Popular tools
    - Ping: RTT and loss
    - Traceroute: path and RTT
  - Problems:
    - Impose extra traffic on network and distort its behavior in the process
    - May impact the behavior of the network (self interfering)

Determining *How* to Measure

- Passive measurements
  - Observing network traffic at the measurement point(s)
    - Packet capture (wireshark)
    - Flow-based measurement tools (routers)
    - SNMP tools (mostly used)
  - Perform analysis for various purposes

- Used to perform various traffic usage/characterization analysis/intrusion detection

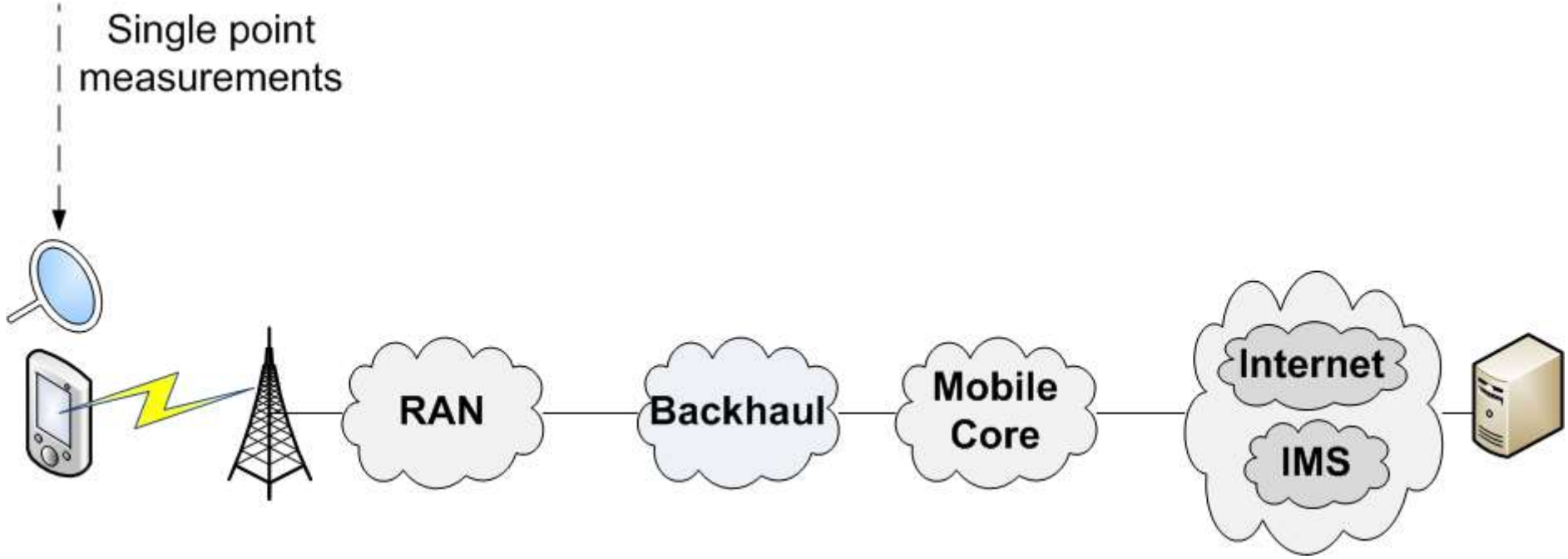
- Problems
  - LOTS of data!
  - Privacy issues
  - Performance issues (wire speed packet capture and analysis)

Comparison of active/passive measurements

	Active measurements	Passive measurements
Configuration	Multi-point	Single or multi-point
Data size	Small	Large
Network overhead	Additional traffic	- Device overhead - No overhead if splitter is used
Purpose	Delay, packet loss, availability	Throughput, traffic patterns, trends, & detection
CPU Requirement	Low to Moderate	High

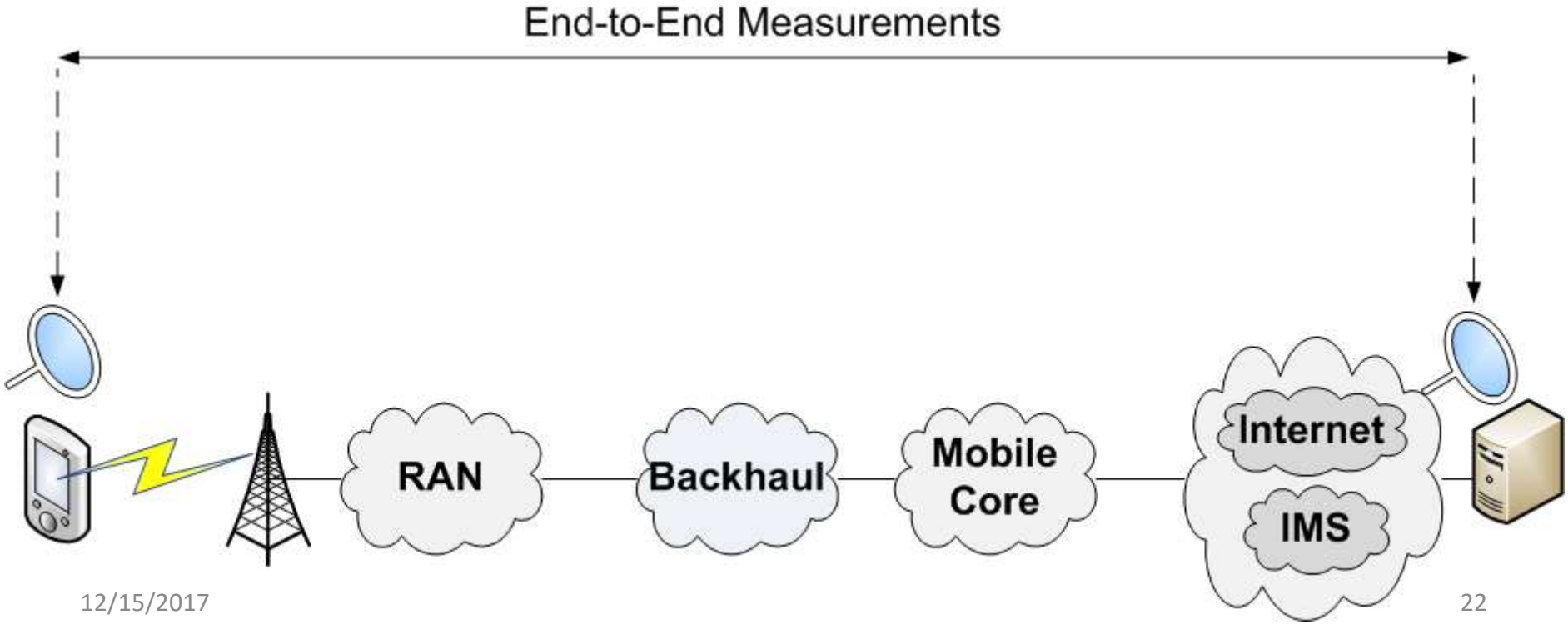
Determining *Where* to Measure

- Single point measurements
  - Provide partial view of the network



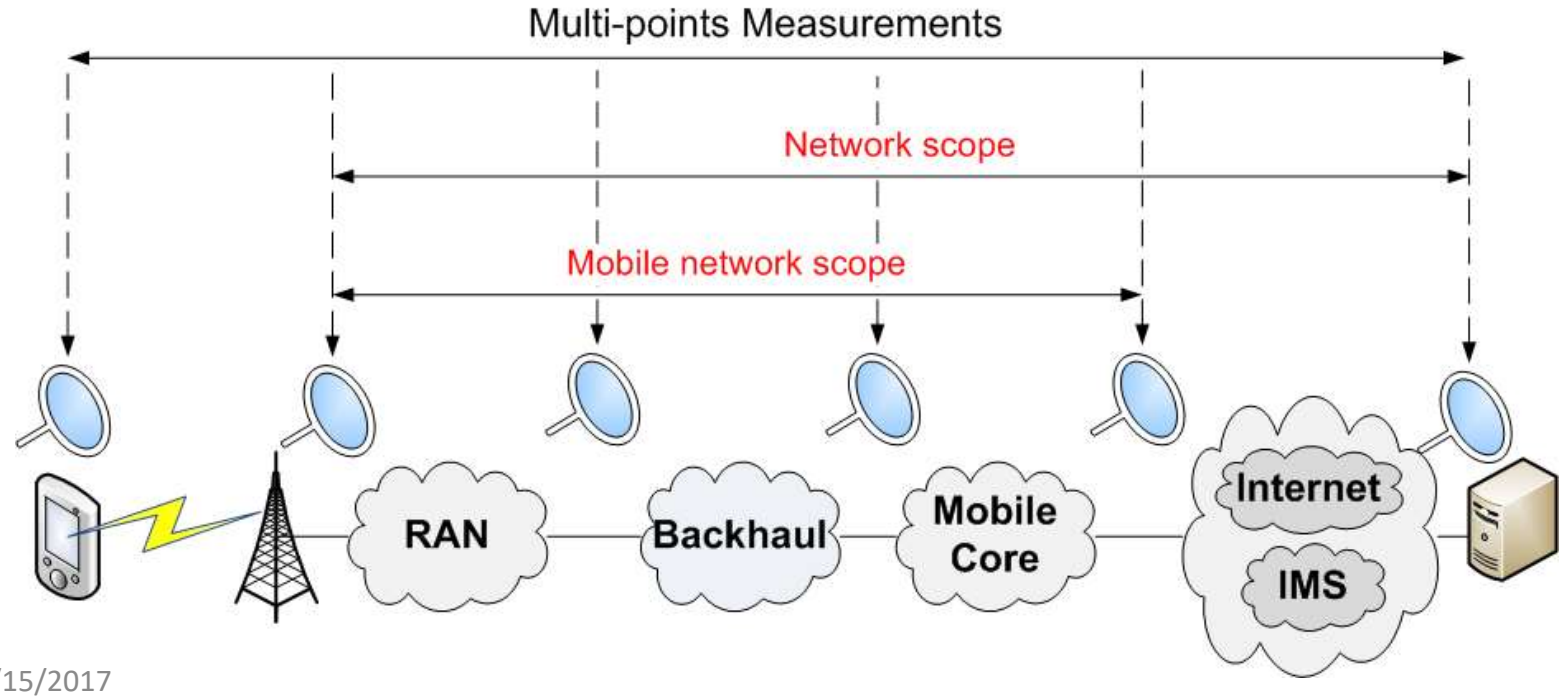
Determining *Where* to Measure

- End-to-end measurements
  - Provide a view on the performance between the the end points



Determining *Where* to Measure

- Multi point measurements
  - Provide a view on the performance in the different "monitored" segments of the network



Seminar plan

- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

Limitation of measurements

- For years, monitoring used primarily
  - Global traffic measures provided by Simple Network Monitoring Protocol
  - Traffic data provided by routers (netflow data)
- These information are always of great interest
- However, are they sufficient?

Limitation of measurements

- What if we need to know:
  - Who's using Skype, P2P, VoD, etc.?
  - Most popular applications on the network
  - Proportion of VoIP calls with bad quality
  - Quality of experience for video streaming
- New means for providing accurate traffic measures is required
  - Deep Packet Inspection is a good candidate

Seminar plan

- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

Deep Packet Inspection

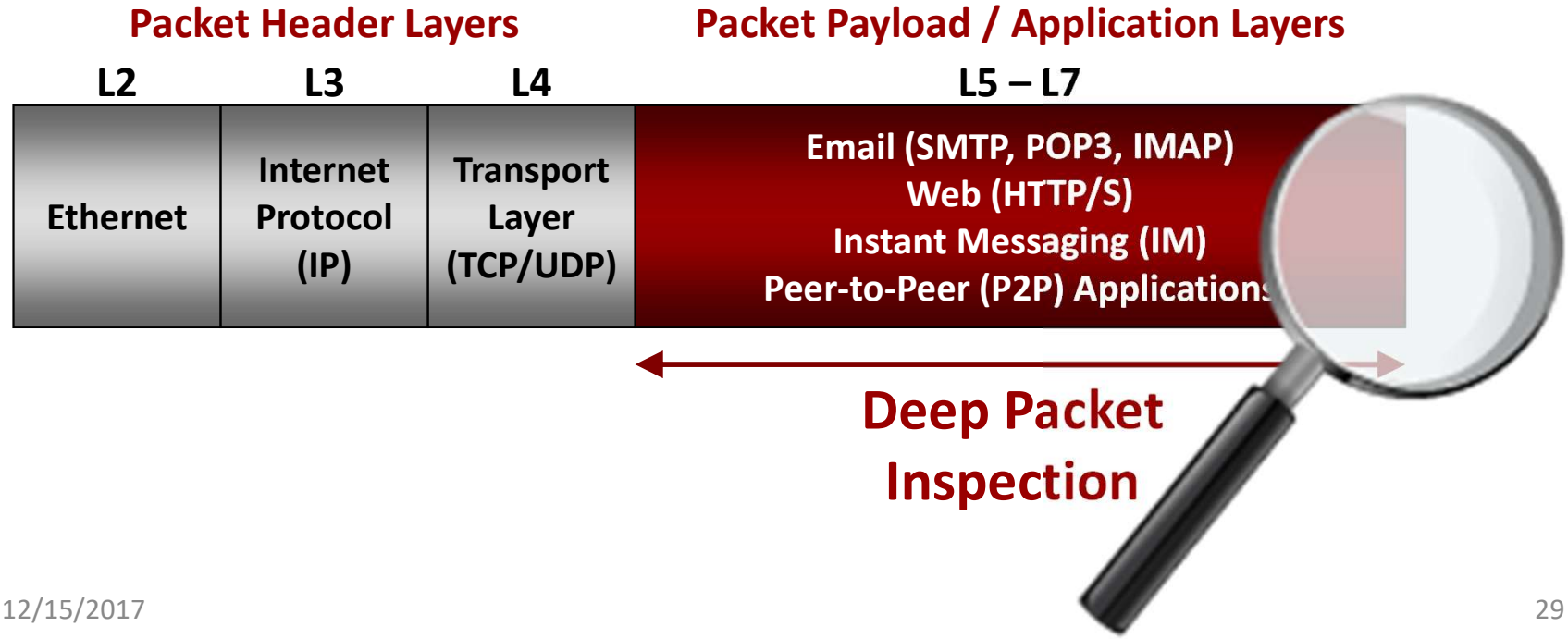
What is DPI and why it is needed

Application classification

Traffic attributes extraction

What is DPI

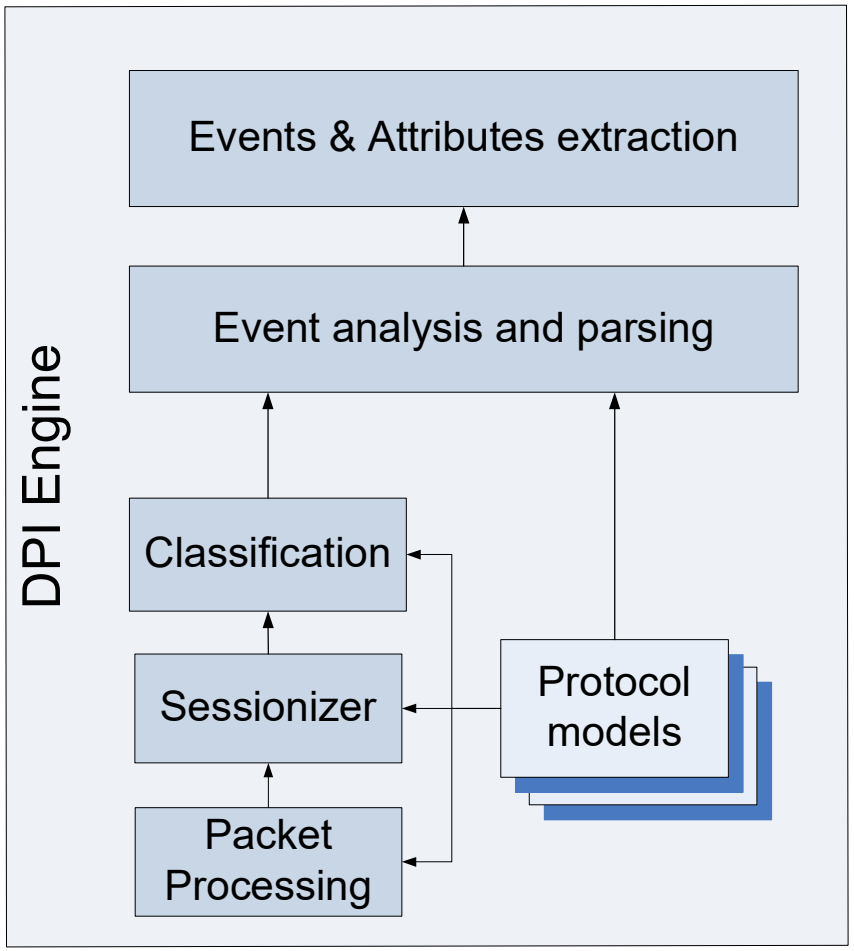
- Technology consisting of digging deep into the packet header and payload to “inspect” encapsulated content
  - Content may be spread over many packets



Why to DPI

- Network Visibility
  - Understand how bandwidth is utilized
    - What is the application mix
    - Who is using what, where and when?
- Traffic Management (Application Control)
  - Block undesired traffic (spam, worms, etc.)
  - Prioritize and shape traffic (limit P2P, QoS, QoE)
  - Advanced policy enforcement
  - Zero Facebook, OTT services, per application policy rules
- Network management
  - Advanced billing (abandoning the unlimited data plans)
  - New pricing may appear soon (user defined preferred applications for free, fees applies for the rest of applications)
- Security
  - Understand network attacks
  - Core component in next generation firewalls
- Etc.

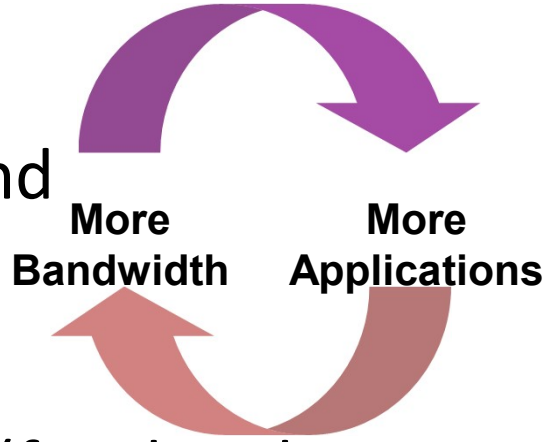
Inside DPI



- Group packets belonging to the same session
- Application classification
  - Detect application type (Skype, Bittorrent, etc.) or application family
  - Considered as the core of DPI
- Protocol decoding and attribute extraction
  - Parse the packet structure (this depends on the protocol & application)
  - Get protocol attributes (IP @, port numbers, …)
  - Get session attributes
  - Events and attributes may involve different packets
    - Attached file of an email

DPI Drivers

- Bandwidth (market driver)
  - High penetration rate for broadband
    - At home: ADSL2+, VDSL2, FTTx
    - Mobile: 3G+ and 4G
  - Bandwidth per user is ramping up (fixed and mobile)



- Legal Interception (legislation driver)
  - Requirement for service providers
  - DPI is a core component

Market drivers: Applications

Video
- Expected to reach 60% of mobile traffic
- Web video (Youtube, daily motion, video channel) + Telcos (IPTV & VoD)

Social
- Increasing traffic with diverse media contents (embedded video)
- High signalling overhead (short frequent sessions)

P2P
- Continue to be highly popular (~ 40-60% BW)
- Use encryption BitTorrent, eMule
- Viable choice for providers (Warner Bros, Spotify)

VoIP
- Skype, GoogleTalk, Yahoo!Voice, Facebook?
- Serious competitor for traditional telephony!

Gaming
- Consoles & PC offer “over the network” gaming experience
- Stringent Bandwidth & Latency requirements

Seminar plan

- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

Application classification: the challenge

- High number of applications and protocols
  - Same Application – Different Implementations/versions
    - Bittorrent has more than 30 different client implementations
    - IM or VoIP don't use similar protocols
  - Evolving Architectures
    - Client/server, Cashes, P2P, Client's network surroundings: Firewall/NAT, Proxy
  - Various Clients: PC, Smartphone, Gaming Console
  - Symmetric vs. Asymmetric
- Frequent Updates
  - Can vary from every year to every month
  - Typically will affect protocol format
- Use of Encryption (Obfuscation)
  - Primarily designed for counter measuring operator's throttling and monitoring efforts (eMule, Bittorrent)
  - In some cases protect proprietary implementation (Skype)
- Need to differentiate use
  - "Good" (legit streaming, SW updates) vs. "Bad" (pirated file sharing) P2P
- Need to recognize application subtleties for proper actions
  - Example: MSN IM – block VoIP & Streaming, allow Chat

Application classification: Techniques

- Classification techniques
  - Port based classification
  - Pattern matching based classification
  - Statistical classification

Application classification Evaluation criteria

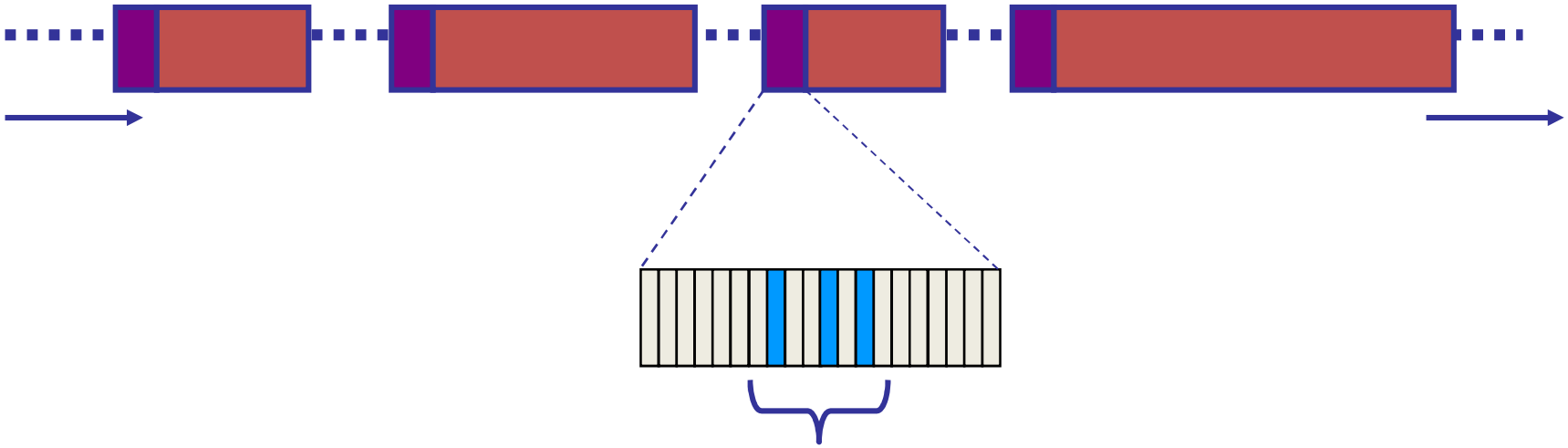
- Completeness
  - Ratio of the application detection count over the expected detection count
  - Low detection completeness indicates many false negatives.
  - A false negative is the inability to classify a flow of application A as a flow of application A.
- Accuracy
  - Ratio of the correct detections over the detection count
  - Measures how correct the detection technique is
  - It may not be more than 100%.
  - The lack of accuracy leads to false positives
  - A false positive is the classification of application B as being application A.

- Impacted if the classification technique is weak
- Target: reduce false positives and false negatives in order to reach a sufficient enough accuracy

Analysis by Port Numbers

- Reasoning:
  - Many applications and protocols have assigned port numbers (and widely used)
  - Example: email
    - Incoming POP3: 110 (995 if using SSL)
    - Outgoing SMTP: 25
- The Good - It's easy ☺ The Bad - It's too easy ☹
  - Many applications disguise the traffic using ports usually used by different protocols (80, 25, 110, …)
    - Firewall and Nat traversal
  - New or unknown protocols
  - Applications that choose a random port number
  - Accuracy ~ 30 – 70 % (close to 0% for some applications: P2P)

Analysis by Port Numbers



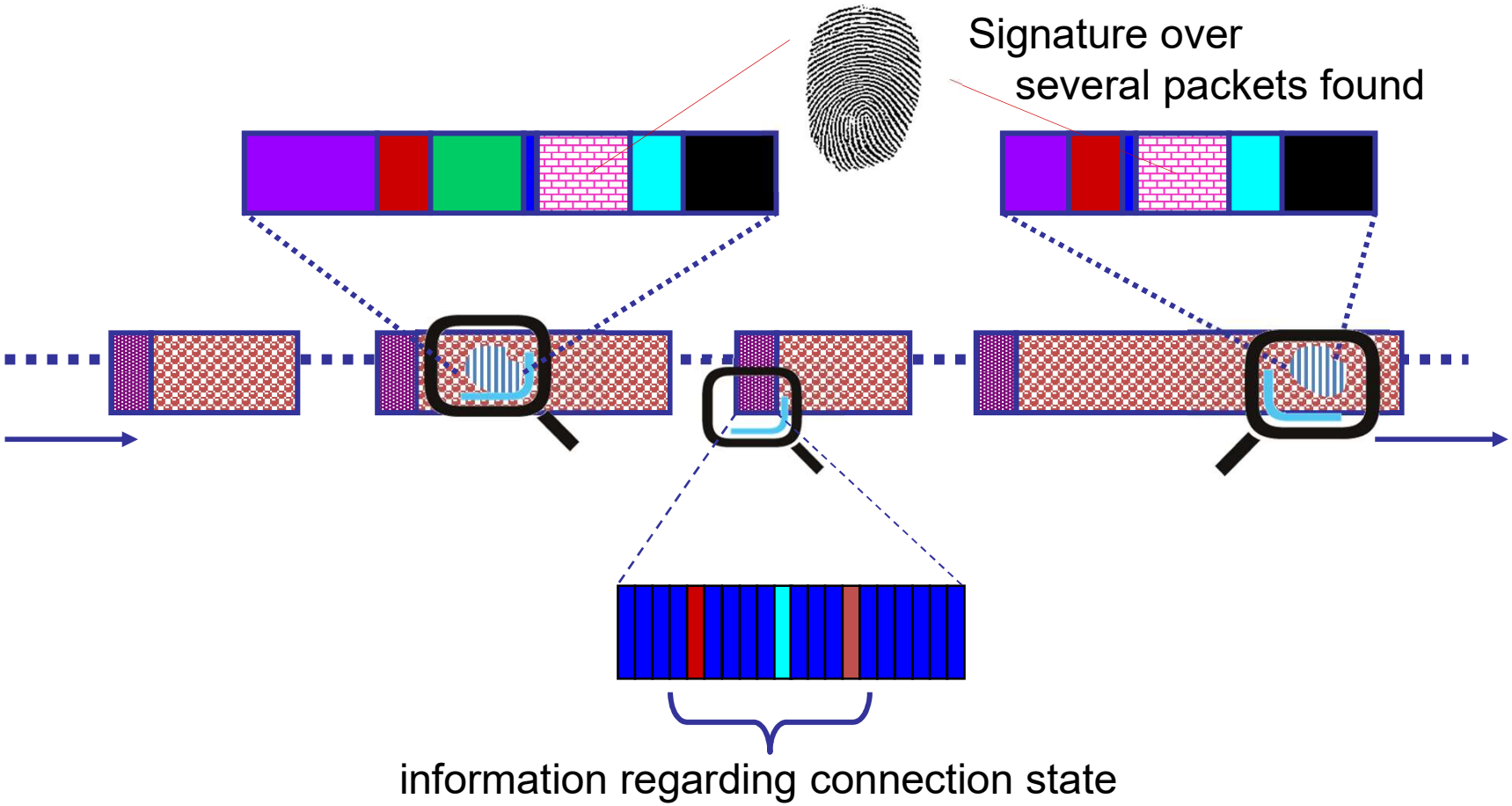
Basic header information (port numbers)

Analysis by Pattern Matching

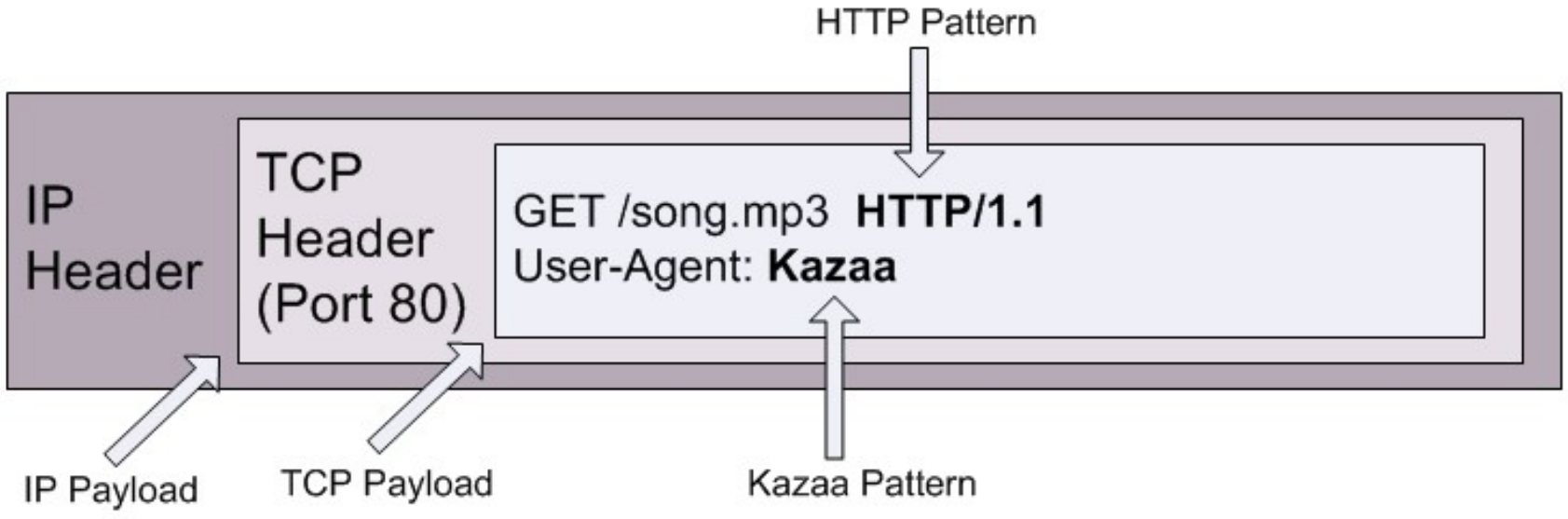
- Reasoning:
  - Many applications have pure textual identifiers
    - Defined in the application or protocol specification documents (RFCs, 3GPP, ITU)
    - Ex: HTTP request must start with

      "**Method** URI **HTTP/1.1**"
- Easy to search for
  - Very easy if in a specific location within a packet
- Uniqueness not always guaranteed
  - Risk to have false positives!
- Pattern may involve different packets
  - Track the connection state and signature matching

Analysis by Pattern Matching



Analysis by Pattern Matching

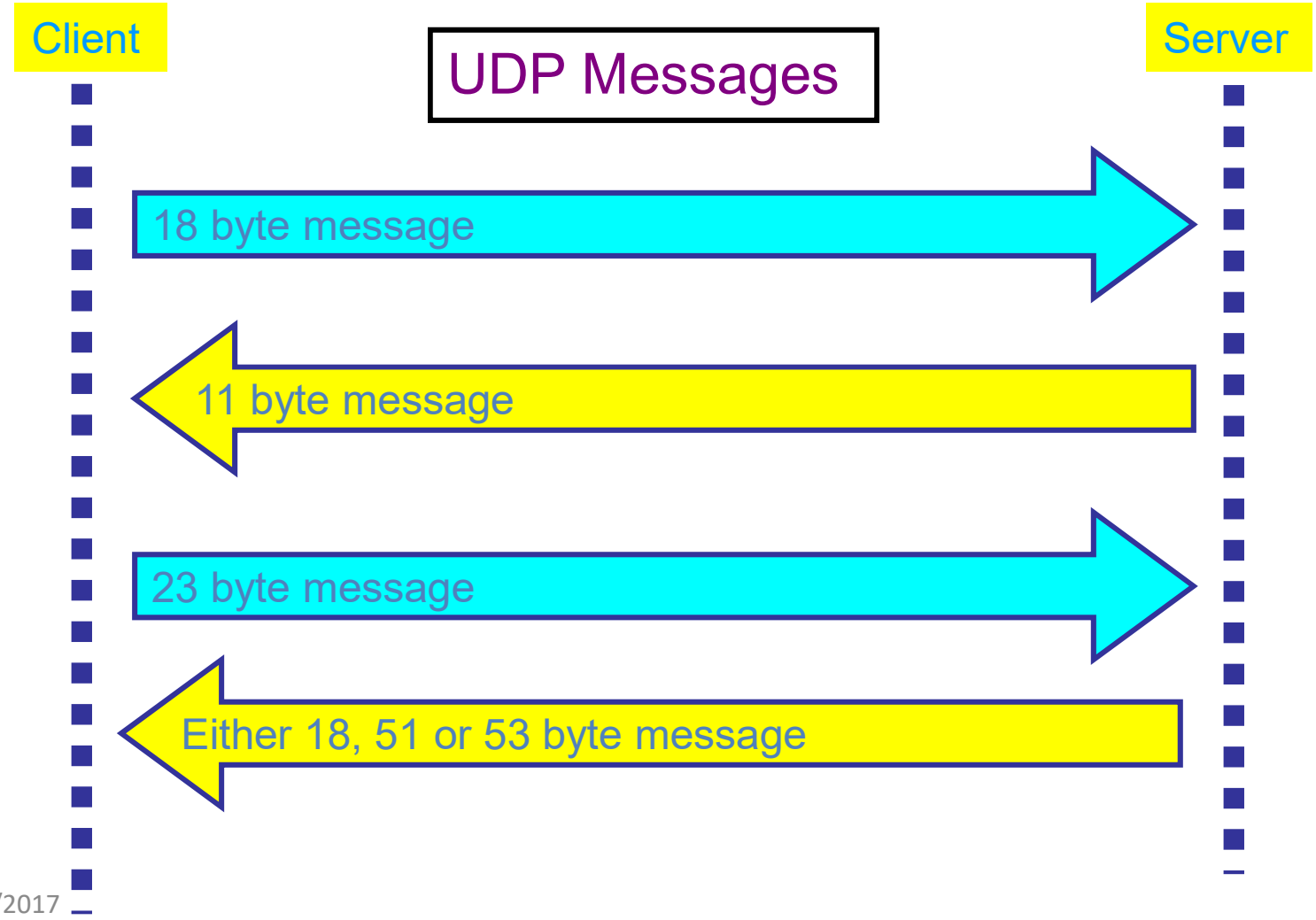


Behavior and statistical analysis

- Many protocols have statistical and behavioral “signatures” that are not related to the data contents:
  - Packet size
  - Inter-arrival delay
  - Specific exchange that can be assimilated to a state machine
- The detection requires a number of packets
- Example
  - Very close inter-arrival delays with low deviation from the average (VoIP)
- Extremely effective analysis when application uses encryption or obfuscation
  - Or simply when access to the payload is not possible
  - Classification in the dark

Behavior and statistical analysis

Skype (Old Version) connection setup



Packet classification: an expensive operation

- High memory and processing requirements
  - Will increase with the number of supported protocols and applications
- Requires dedicated high capability hardware
  - Multi-core technology
  - Line rate packet capture capabilities
- How it can be optimized???

Packet Classification: Optimization perspectives (1)

- Observation: traffic flows with the same server IP@ and port number will most probably have the same application type
- Idea:
  - Cache server identifiers (IP@ and port number) along with the application class/type
- Work mostly for client/server based applications
  - P2P traffic still needs to be classified
  - Overall gain though
- Preliminary evaluation shows a global gain of 30 ~ 60 % (vary with the application mix)
  - Looks very promising
  - Results to be confirmed on large scale evaluation

Packet Classification: Optimization perspectives (2)

- Observation: If a user is currently using an application, he will most probably use it again in the very near future
  - Take yourself as examples ☺
- Idea:
  - Prioritize the classification by checking first the protocols and applications the user has recently used
    - Require to maintain a cache per user (few items 5-10)
- Why not to combine both ideas
  - Cache of server identifiers and application types
  - Per user cache of recently used applications

Seminar plan

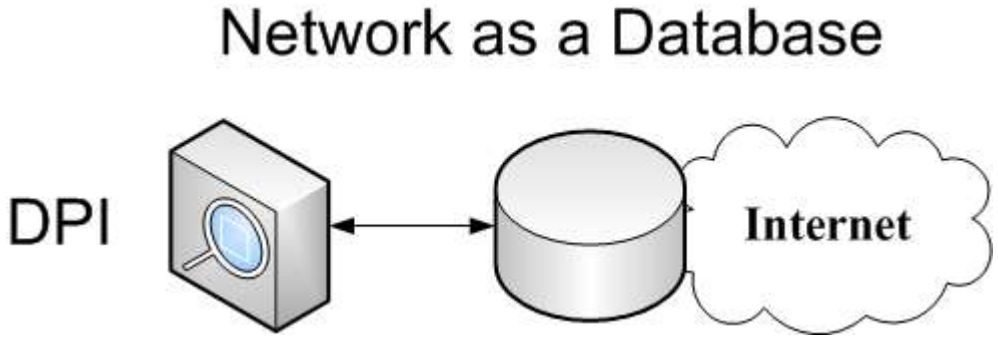
- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

From classification to attributes and events extraction

- Application classification is a first step towards accurate traffic information extraction
  - How can we get the HTTP method (Get, Post, etc.) if we don't know the type of the traffic
  - When the application type is known decoding becomes "easy"
- What are traffic attributes?
  - Protocol field derived from the packet data: IP@, attachment size, encoding type, etc.
  - Flow parameter: packet mean size, inter-arrival delay, packets lost, reordered, etc.
  - The application class can be considered as a traffic attribute

Attribute extraction with DPI

- With the extraction capability, DPI can provide input for other mechanisms as:
  - Security analysis
  - User quality analysis
- Imagine the network as a database and DPI as an engine to extract data from this database!



Network as a Database

- Select user_id, perceived_quality Where (application = Video AND protocol = RTP)

- Select flow_id Where (application = email AND attachment is executable)

Seminar plan

- Network monitoring
  - Needs for network monitoring
  - Measurements (what, where and how)
  - Limitations (briefly)
- Deep Packet Inspection
  - What is DPI and why it is needed
  - Application classification
  - Traffic attributes extraction
- Security monitoring with DPI
  - Abstract description
  - Challenges
  - Security properties

Security Monitoring with DPI

Abstract description

Challenges

Security properties

The HBGary Hack

- HBGary - experts in computer security
  - computer forensics and malware analysis tools to enable the detection, isolation, and analysis of worms, viruses, and trojans
  - implementing intrusion detection systems and secure networking
  - performs vulnerability assessment and penetration testing of systems and software
  - rootkit.com is a respected resource for discussion and analysis of rootkits
- CEO Aaron Barr wanted to unmask Anonymous
  - Those responsible for co-ordinating the group's actions, including the denial-of-service attacks that hit MasterCard, Visa, and others late last year
- Anonymous response
  - HBGary's servers were broken into, its e-mails pillaged and published to the world, its data destroyed, and its website defaced. As an added bonus, a second site owned and operated by Greg Hoglund, owner of HBGary, was taken offline and the user registration database published

(source) http://arstechnica.com/tech-policy/news/2011/02/anonymous-speaks-the-inside-story-of-the-hbgary-hack.ars/1

The HBGary Hack

	Attack Method	Attacked System	Vulnerability	Lost Assets
1	**SQL Injection**	CMS on HBGary Federal's website, hbgaryfederal.com	CMS with missing validity check of SQL Parameters	usernames, e-mail @ & password hashes
2	Password cracking using rainbow tables	Password hashes from 1	Hashes without salt, weak passwords	clear text passwords
3	Unauthorized use of passwords from 2	E-mail, Twitter accounts, and LinkedIn accounts of HBGarry officials	Password double use	Email accounts of HBGary officials
4	Unauthorized use of passwords from 2	Machine running support.hbgary.com	Password double use	Non-superuser account of HBGary official
5	Privilege escalation	Machine running support.hbgary.com	Privilege escalation vulnerability, system not up to date	Full access to HBGary's system, gigabytes of backups and research data
6	Social engineering	Machine running rootkit.com		Integrity of rootkit.com

The HBGary Hack: Where can security monitoring help?

	Attack Method	Attacked System	Vulnerability	Lost Assets
1	**SQL Injection**	CMS on HBGary Federal's website, hbgaryfederal.com	CMS with missing validity check of SQL Parameters	usernames, e-mail @ & password hashes
2	Password cracking	Password hashes from 1	Hashes without salt	clear text passwords
3				
4				
5			up-to-date	backups and research data
6	Social engineering	Machine running rootkit.com		Integrity of rootkit.com

SQL injection: In top ten most known vulnerabilities
(Top ten vulnerabilities are responsible for 60% of software bugs!)

Multi-lines security system: Is it possible to detect security attacks (SQL injection in this case) by inspecting the data encoded in inbound traffic?

Security monitoring with DPI: Abstract description

- The concept:
  - Detect the occurrence of **events** on the network
    - Input provided by DPI
    - Event can be: packet arrival, HTTP POST request, etc.
  - Inspect and analyze the succession of events to detect **properties**
    - Property: Succession of events that are linked with “time” and “logical” constraints
      - If we detect event “A”, then we MUST detect event “B” before 10 seconds
- The idea:
  - Monitor the network looking for the occurrence of properties.

Security monitoring using DPI: Abstract description

- Example: SQL injection
  - www.abcd.com/page?name=**Select * Where 1**
  - The events events:
    - HTTP GET request
    - URL parameter contains SQL statement
  - The property
    - It is **not allowed** to have a URL parameter containing SQL statement in an HTTP GET request
  - If the property is detected on the network then most probably there is an attack attempt!
- Nice Theory! But very challenging

Security monitoring using DPI

- Challenges
  - The number of events that can occur on a network is huge!
    - Solution: Use DPI for the events extraction
    - Group events/attributes by application and add new ones when needed
  - The expressivity of the properties (need to combine time and logical constraints)
  - Complex analysis and processing especially in high bandwidth links
    - Optimization techniques, multi-core implementations, smart traffic filtering

Properties Expressivity

Considering security monitoring, properties can be used to express:

- A Security rule describes the expected behavior of the application or protocol under-test.
  - The non-respect of the Security property indicates an abnormal behavior.
  - Set of properties specifying constraints on the message exchange
    - i.e. the access to a specific service must always be preceded by an authentication phase
- An Attack describes a malicious behavior whether it is an attack model, a vulnerability or a misbehavior.
  - The respect of the Security property indicates the detection of an abnormal behavior that might indicate the occurrence of an attack.
  - Set of properties referring to a vulnerability or to an attack
    - A big number of requests from the same user in a limited period can be considered as a behavioral attack

Properties Expressivity

- A security property is composed of 2 parts:
  - A Context
  - A condition to verify

- The “Context” and “condition” of a property are composed of:
  - Simple events
    - Conditions on attributes (IP @ equal to 1.2.3.4)
  - Complex events linked by
    - Logical operators (AND/OR/NOT)
    - Chronological operator (AFTER/BEFORE)

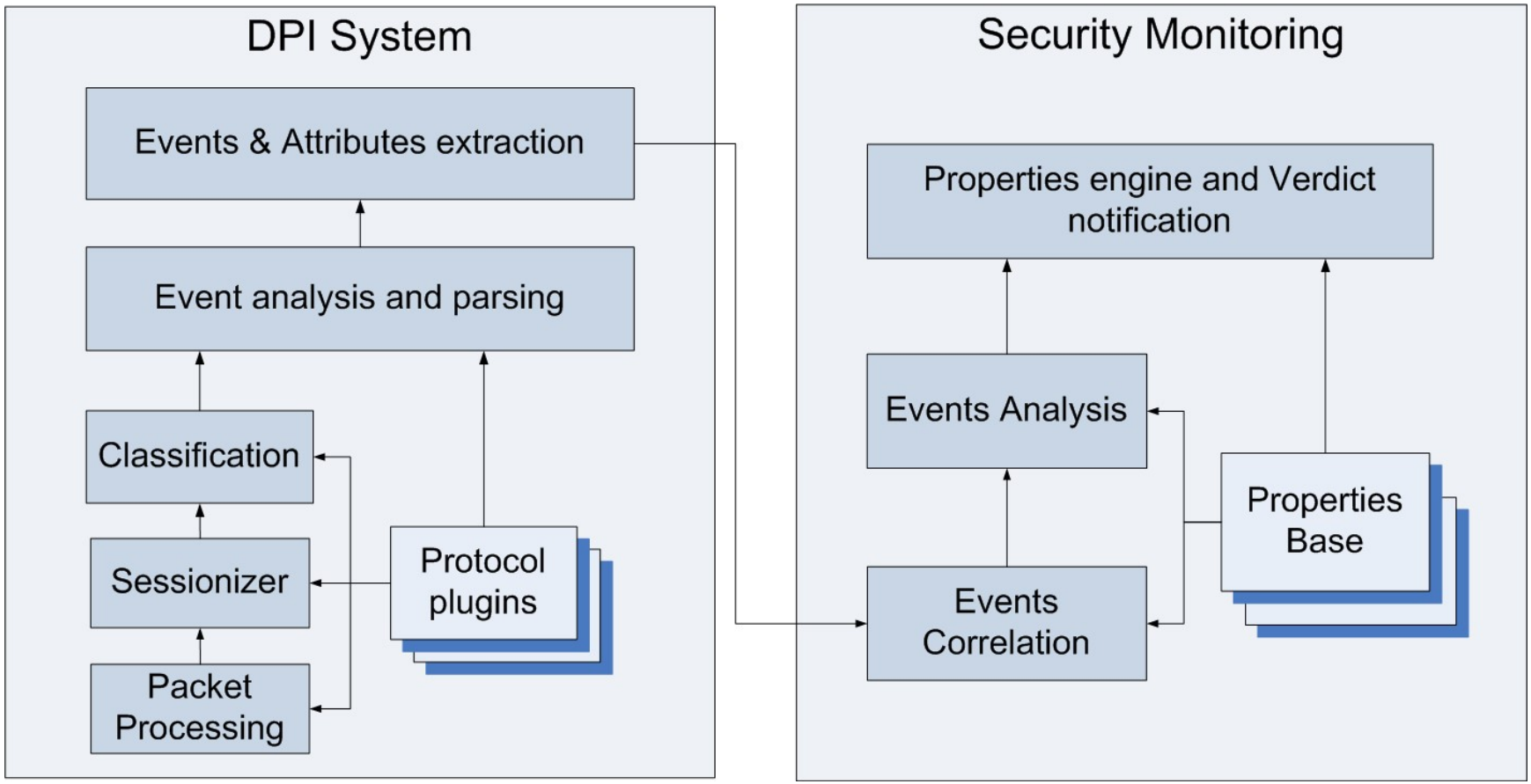
Properties Expressivity

- A security property is composed of 2 parts:
  - A Context
  - A condition to verify

If an HTTP Response is received (this is the context)

Then an HTTP request should have been received **before** (condition to verify)

Security monitoring with DPI



THANK YOU

Wissam Mallouli
wissam.mallouli@montimage.com

- shawez@jnu.ac.in
- Lucie.poggi@agroparistech.fr
- Mohamed-reda.belbahri@espci.fr
- mariabargues1@gmail.com