ITU Workshop on "ICT Security Standardization for Developing Countries"

(Geneva, Switzerland, 15-16 September 2014)

Validation and Reverse Business Process Documentation of on line services

Maurizio Talamo,

Full Professor, University of Rome "Tor Vergata",

maurizio.talamo@uniroma2.it

Business problem overview

- Problem owner
  - Organizations that provide informational, administrative and economical services over the Internet
- Issues
  - Business intelligence and analytics
    - "on line" acquisition and "real time" integration and processing of information from the services issued over the Internet
  - Decision making
    - improve the decision making activities through "real time" knowledge of the behavior/status of the services issued over the internet
  - How to effectively implement and use "process mining" and "online" data acquisition techniques

Given a web-service issued over the internet extract process "logs" directly from the network data

- Data
  - The elementary items passing through the network, at different level of complexity
- Services/Security
  - Specific reconstruction activites, starting from the acquired data
- Transactions
  - Specific discovery activities of the relationships between services issued, where a transaction can be viewed as a "composition/sequence of single services"

The Data

- The elementary items passing through the network, at different level of complexity:
  - Low level communication protocol (MAC address, TCP/IP address and port)
  - High level communication protocol (HTTP host and headers)
  - Application protocol (HTTP URL and Parameters, HTML page items, …)
  - Communication errors/anomalies

Services/Security

- As a result of a specific reconstruction activity, starting from the acquired data it is possible to obtain
  - Numerosity of the services issued
  - Quality of services: Response time of the services, data loss, usage statistics of the services issued including user sessions timing, discovery of critical services, discovery of the most/less used services
  - Numerosity of the users who access to the services
  - Security level (security protocols used, controlled access, dynamic pages usage, …)
  - Discovery of DoS attacks of different nature (at tcp level, at http level, fast repetition of patterns, upload of too big files, access to too big URLs, inadequate number of parameters, …)
  - Territorial distribution of the traffic, too big files and data, usage of old communication protocols and languages (HTTP/1.0, HTML 4, XACML 2.0, …)

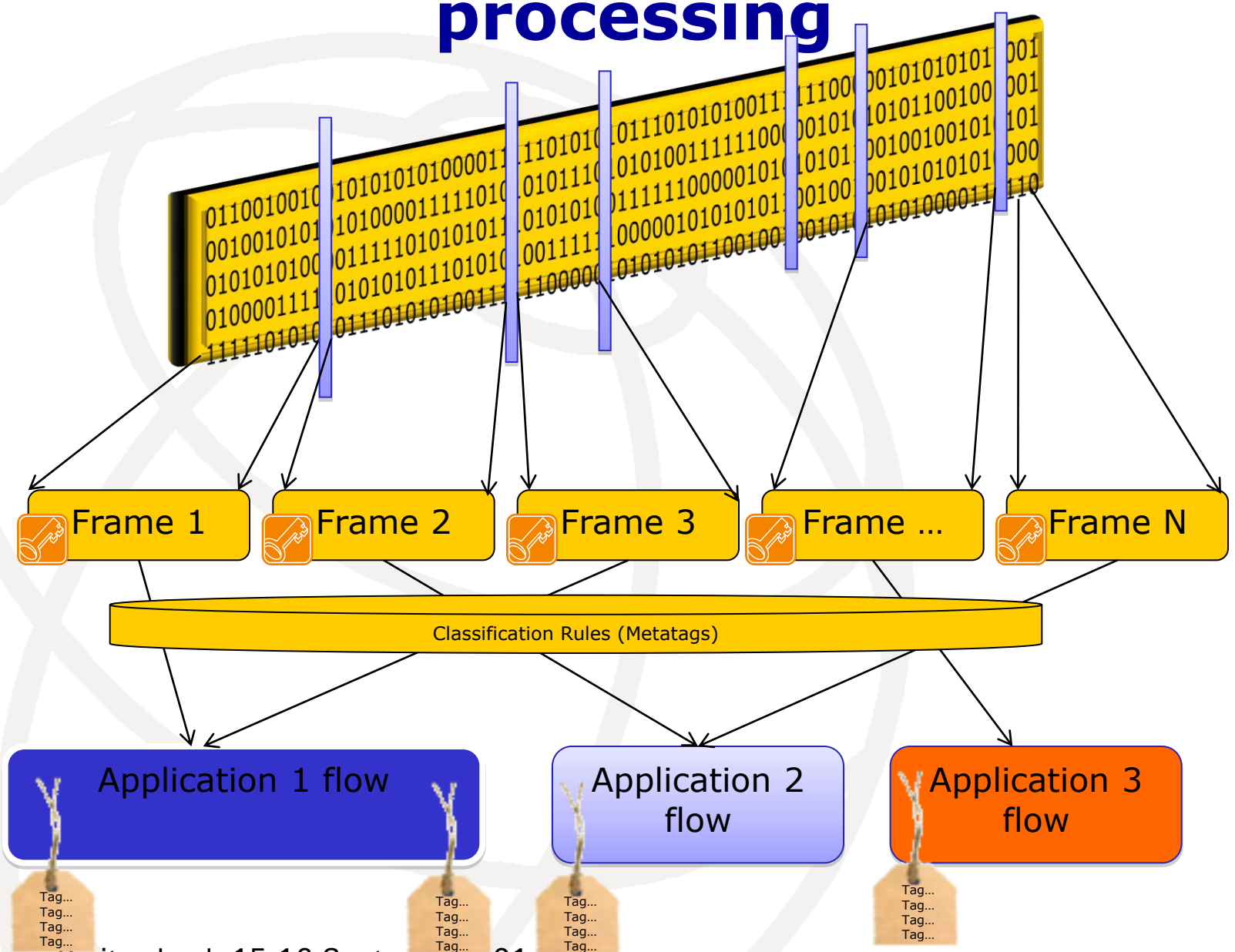
Transactions

- As a result of a specific discovery activity of relationships among services issued, where a transaction is a "composition/sequence of single services"
  - Certification of the correctly terminated transactions (made by one or more services)
  - Certification of unfinished transactions (when and at which step the transaction has stopped, the user, …)
  - Anomalous behaviours (transactions/services that do no comply with the designed behaviour)
  - Complex security attacks (phishing, identity theft, privilege escalation, …)
  - Archiving of the original logs in an "untouched" way for auditing, ensuring that one can only access data for which access has been authorized.

Some technical issues

- How to manage a large quantity of (real time) logs and how to discover relationships among them.
- How to enable a fast search (real time) on the logs without modifying them.
- How to generate patterns and recognize them on the information flow, to be evaluated as anomalies or not.
- How to maintain original logs in an "untouched" way for auditing, but, at the same time, build indexes to search them and to avoid exposure to unauthorized users.
- How to avoid data loss when acquiring logs "directly" from the networks.
- How to permit a search of the logs based on complex keywords and structured descriptions of properties where temporal order is important.
- To query and recognize chains of events occurring at different locations and to relate them using tags directly acquired "online".
- To discover anomalies and errate behavior of the services issued over the internet caused by the customer.
- To support the technical management of the sites of the customer, including most used pages, IP traffic, territorial distribution of the traffic, too big files and data, usage of old communication protocols or run time libraries, …
- To support the security of the sites, discovering DoS attacks of different nature (at tcp level, at http level, fast repetition of patterns, upload of too big files, access to too big URLs, inadequate number of parameters, …)

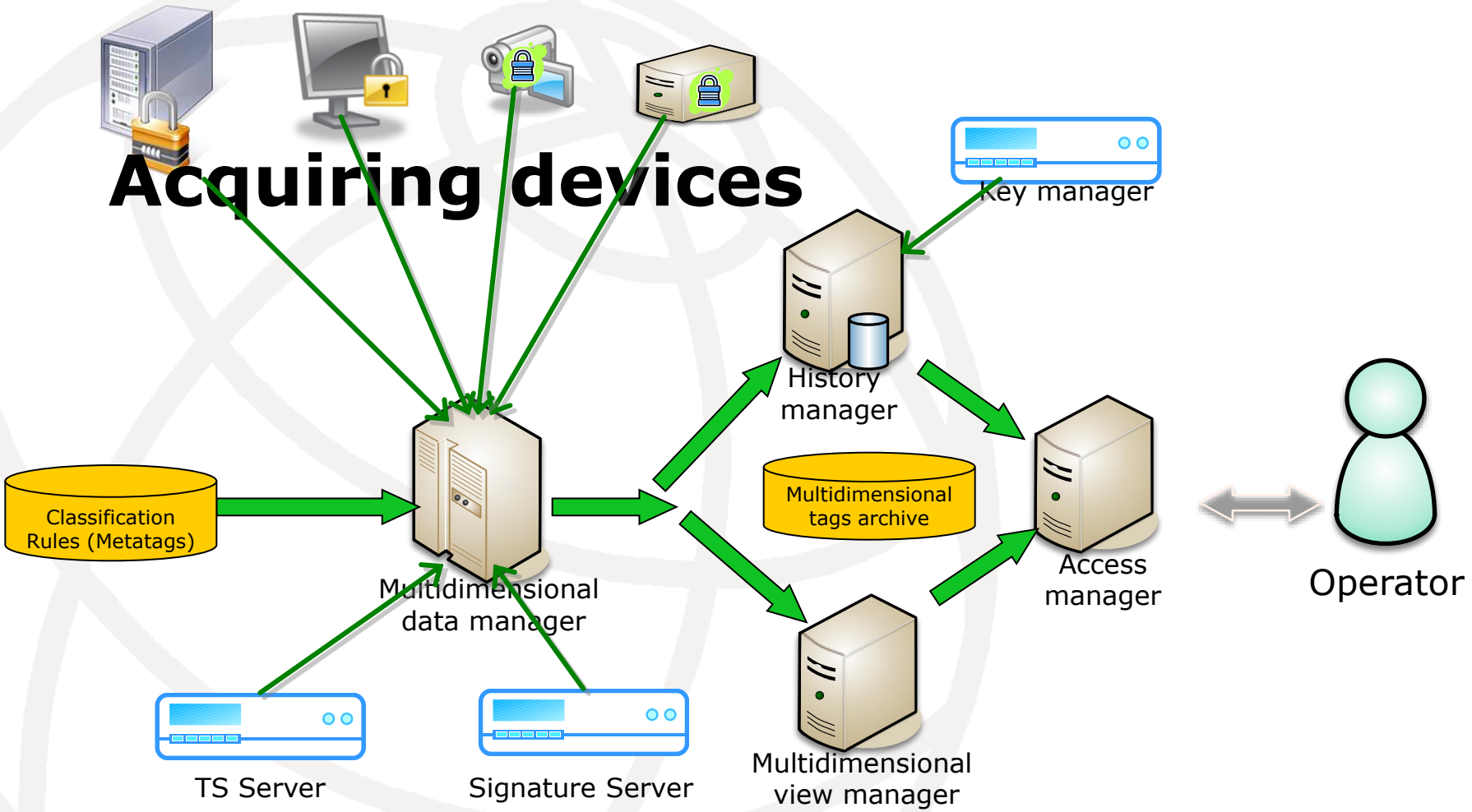
On line data acquisition and processing



Basic functional requirements

- To acquire logs directly from the network the acquiring devices should
  - Automatically reconstruct original applicaton flows from the data frames acquired on the network
  - Sign digitally the original data frames and associate them to the reconstructed application flows, making them usable as proof in criminal or administrative proceedings
  - Associate tags, from a set of selected Metatags, to the entire or parts of the reconstructed application flows
  - Provide a query language for selecting all the reconstructed flows and the original signed and unmodified network frames, based of the associated tags or temporal sequences of tags

Modular architecture



Modular architecture

- The raw data are aggregated from and classified according to the "Classification Rules" in the system "Multidimensional Data Manager" (MDM)
- The MDM locates in real time the MetaTags in the raw data (defined by the rules of classification), structures them according to a scheme of multidimensional access and associates them with the flow of data related to the process instances
- The MDM builds and maintains also a multidimensional data structure containing all the meta tags and associated tags used for classification of the application flows
- The output data from MDM reaches the History Manager, which stores and encrypts all elementary process flows.
- The History Manager provides on demand the keys, managed by the Key Manager, to decipher the individual elementary frames and produces a "log" of such requests.
- The tags, organized in multidimensional access structure, are directly usable, through the query language of the View Manager, in order to be queried and to provide the original frames or the resulting classified application flows to the user.