

**INNOVATION REPORT**

# Processing massive data flows in real time enables novel financial and communications services



Project leader: Pedro Soria-Rodriguez (ATOS Origin)

*The PubSub4RT project has defined and developed an infrastructure to handle massive data flows in real time using publish/subscribe technology. This makes it possible to provide novel services based on processing these flows, particularly in the financial, business intelligence and telecommunications domains. PubSub4RT middleware can also help solve the usual security problems facing financial entities and Internet service providers – such as denial-of-service attacks, spamming and phishing. The resulting platform is highly scalable and has been successfully demonstrated on wide-area networks for credit-card fraud detection.*

There are many applications in telecommunications, banking, security, Internet services and sensor networks which require the treatment of massive amounts of flowing data and often over wide geographic areas. Examples include anti-spam and anti-virus filters for multinational company email systems, processing the output of large sensor networks, protection against distributed denial-of-service attacks, anti-fraud services for global credit-card payments, financial services over stock quotes, real-time processing of call description records to fight fraud in mobile telecommunications and real-time processing to enable targeted business intelligence.

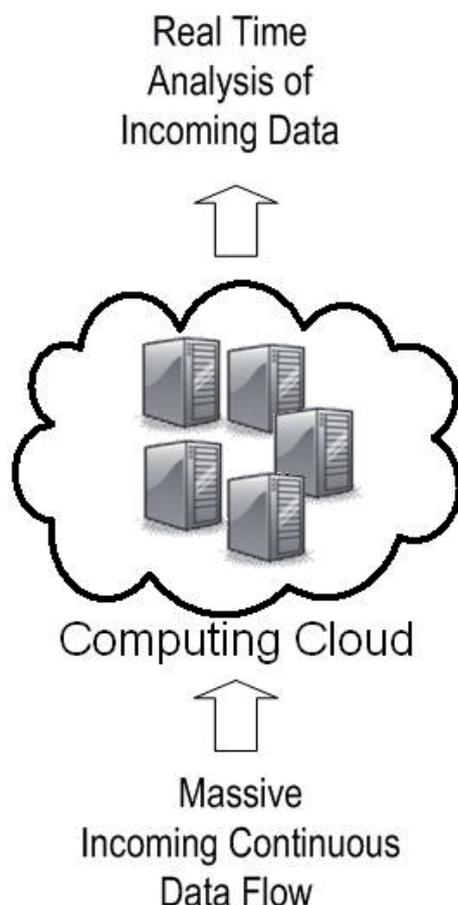
Until now, data-streaming processing infrastructures have had only modest scalability, limiting their range of applications. These infrastructures are characterised by having to process the whole information flow in each node, which then become individual bottlenecks.

PUBSUB4RT has developed an architecture that enables information flows to be processed in parallel between a subset of nodes, avoiding concentrating the information flow in a single node. This distributed processing approach results in a highly scalable infrastructure able to tackle a new class of applications not possible with the current technologies.

### Two-tier approach

At the heart of the PubSub4RT approach is a two-tier middleware infrastructure. A central core with a data-streaming engine makes it possible to handle massive data flows in real time, while an additional upper layer provides publish/subscribe on-line services to the client applications – for example subscribing to events such as notifications of financial market conditions in business intelligence applications.

A major novelty is the ability of the platform to support services both on a single domain and in multiple domains separated by a wide area network (WAN). Thus the market relevance is particularly high, including in telecommunications, banking and financial services, network security, Internet service provider (ISP) services and sensor networks such as those used to



## INNOVATION REPORT

monitor road traffic, the quality of running water in large cities or environmental conditions.

The parallel data-streaming system enables aggregation of hundreds of nodes to process a single query and handle a million events a second. The platform has also been enriched with two self-management features:

1. **Self-optimisation** – dynamic load balancing makes it possible to maximise the throughput for a given configuration; and
2. **Self-provisioning** – new nodes are provisioned and decommissioned to minimise the resources required to process the incoming load.

In addition, the data-streaming middleware provides a SQL-like language to handle queries. This offers both 'stateless' operators, which provide output based on each individual input string, and 'stateful' operators, which provide output based on a sliding time window.

Special attention was paid to the network platform as a high fraction of the overhead in these types of application is network processing. This network overhead was reduced by cutting the number of copies made in the network path, introducing flow control and batching of the data. The result is a highly scalable data-streaming platform which enables events to be processed at very high rates.

### Credit-card fraud detection

The PubSub4RT approach was applied successfully to a credit-card fraud detection application. Current systems are only able to detect fraud after the occurrence – and this can be hours or even days after such a fraud has been committed with corresponding financial loss. PubSub4RT enables real-time detection of fraud before payment authorisation, avoiding losses.

Moreover, PubSub4RT can offer a choice of two types of query, either: event-based – depending solely on the observed data stream; or profile-based – taking into account both the observed data stream and profiles built on customers' and merchants' normal use.

The resulting platform opens up a wide range of potential applications domains based on a new breed of on-line services:

- **Banking:** anti-fraud, anti-money laundering and market surveillance;
- **Financial services:** buying/selling share options in the financial sector requires processing simultaneously information from thousands of markets to automate financial operations and optimise profit margins in such operations

- **Telecommunications:** anti-fraud operations in mobile phone networks – detection of subscriber identity module (SIM) card cloning has until now been too complex and demanding to attempt – PubSub4RT infrastructure makes it possible to process all mobile phone calls in real time to make temporary and geographically correlations to detect fraud patterns;
- **ISP services:** particularly anti-spam and anti-virus filters – it is much easier to detect spamming patterns over the whole network traffic rather than in individual customer data flows;
- **Network security:** detection and mitigation of distributed denial-of-service attacks as well as general network monitoring; and
- **Business intelligence:** replacing specific off-line processing of massive amounts of information using data-warehousing techniques with real-time approaches that permit fast detection of market trends and targeted offers making use of the information gained.

Other possible applications include: air, road and sea traffic control; border surveillance; environmental monitoring; e-health systems; and cloud computing.

### Exploitation already in progress

Atos Origin, through its Atos Information Security Services (AISS) offering, is already deploying a first test version of PubSub4RT in its business risk and information security activities. A particular application is to provide security in the ICT services management of future Olympic Games as part of its event management activity.

Spanish partner Universidad Politecnica de Madrid (UPM) has been approaching telecommunications operators to identify potential applications of the data-streaming layer developed in the project. UPM has also filed a patent on the parallel data streaming engine in the USA. It already has a contract with Ericsson to explore the data-streaming technology for deep packet inspection and is providing training in the data-streaming area with a course at Ericsson. UPM has also been in discussions with Telefonica Soluciones to explore the application of the data-streaming technology for real time business intelligence.

More information: <http://pubsub4rt.atosorigin.es/en>