Proposition de recherche doctorale

Foundations of an Internet measurement science

Mots clés :

● Directeur de thèse : DARIO ROSSI
● Co-encadrant(s) :
● Unité de recherche : Laboratoire Traitement et Communication de l'Information
● Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
● Domaine scientifique principal: Divers

Résumé du projet de recherche (Langue 1)
Résumé du projet de recherche (Langue 2)

Internet measurements are often considered as a non scientific discipline, though, implicitly, this is already the case since the so called “passive techniques” (where Internet traffic is merely observed) remind the induction and deduction methods of Aristotle, while “active techniques” (where a reaction to a perturbation actively induced on the Internet is the observable of interest) are reminiscent of the experimental method of Galileo Galilei. Yet, a major drawback currently affecting active experiments is their lack of reproducibility, which threatens their relevance and reliability; conversely, passive techniques are often applied to few measurement points and for a short duration, which limits the significance and scope of the deductions. The goal of this thesis is to introduce a rigorous scientific method in Internet measurements, by continuously repeat, integrate and correlate data from different Internet “sensors”, and to systematically repeat the analysis that are now done sporadically at best (PhD Thesis description) The ubiquity of Internet access, and the wide variety of Internet-enabled devices and applications, have made the Internet a principal pillar of the Information Society. Decentralized and diverse, the Internet is resilient and universal. However, its distributed nature leads to operational brittleness and difficulty in identifying and tracking the root causes of performance and availability issues. The first step to improve this situation is Internet measurement: illuminating the currently obscure dynamics of the Internet. Generally speaking, networks have been built conceptually around the pillars of data and control planes, respectively to move data around and to instruct the network on how to precisely move this data. However, Internet architects had not foreseen its growth, and have therefore neglected the need for yet another crucial plane: the measurement plane. The role of this (non yet-existing) plane would be observe, measure and quantify the behavior of the network. Such actions are not only necessary to ensure the proper working of control and data planes, but also and especially to pinpoint the occurrence of performance problems, and possibly assist in determining the Internet path performance throughout the network. In order to perform measurements. Perhaps the most commonplace passive tools are tcpdump or Wireshark and NetFlow designed to let researchers analyse the captured packets or flows, respectively. Beyond these, there are a large number of tools that designed to perform passive measurements, e.g., PF_RING[T3], Tstat[T4] and the CoralReef suite[T5]. Some tools explicitly target specific scenarios (the Aqualab project[T6] to monitor BitTorrent performance to check for changes that might indicate problems with the network), or focus on detecting network neutrality violations (Web Tripwires[T7] detect changes to web pages at some intermediary node; while NANO[T8], Netalyzr[T9] and Glasnost[T10] analyse various properties of an Internet connection, including blocking or shaping of specific services). An equally large number of active measurement tools exist. Several of these have the aim of mapping the structure of the network, attempting to discover the topology of the Internet at a certain level in the protocol stack, usually IP or above. Operators use the results of such measurements in routing and network path optimization. The most obvious example in this area is the ubiquitous traceroute utility, which attempts to trace all IP hops between a source and a destination. iPlane[T11] is a system that provides Internet path performance measurement by extracting an annotated map of the Internet, with measurement points deployed on PlanetLab. Merlin[T12] is a hybrid Internet topology discovery tool at the router level that mixes IGMP probing and traceroute measurements. NetViews[T13] is an end-host (software-based) tool that monitors both control plane and forwarding plane paths on the Internet by leveraging a real time experimental BGP collector. ASP[T14] is a passive monitoring toolkit focusing on providing an Internet map at the AS path level. Other active measurement tools concentrate on measuring various network layer metrics such as delay, jitter, throughput, and packet loss, among others. Beginning with the commonplace but simplistic ping utility, packet-pairs based tools (e.g., capprobe[T15] and pathchar[T16]) focus on measuring delay and performing capacity estimation. Recently, a number of tools target measuring ISP performance, including Dasu[T17] (which concentrates on BitTorrent performance), BISMark[T18] and SamKnows[T19] (which deploy hardware measurement probes in customer’s home networks). All described tools are not programmable, that is, their functionality cannot be easily changed once deployed, and they do not provide easy means for users to improve or change their functionality beyond a narrow set of runtime parameters. Only a few programmable tools/probes do exist. CoMo[T20] is a system for fast prototyping network data monitoring applications by providing a programmable framework. It is oriented towards passive monitoring, thus limiting its flexibility, and it provides raw data as output. No coordination among COMO probes is indeed envisioned. Hence, there are a large variety of existing tools, both for passive and active measurements, yet few on their own are flexible enough to serve as the basis of an open, general and programmable measurement plane for the Internet. This thesis will go beyond the state of art by designing a measurement plane coordinating and integrating passive and active tools. Moreover, by integrating passive and active methods, it will be possible to conceive and develop completely novel hybrid approaches, where probe traffic designed and optimized for measurement purposes is artificially injected (as in active measurements) and then captured passively at various monitoring points as it traverses the network. To the best of our knowledge, the integration and hybridation of passive and active techniques has not been explored so far by the research community. In summary, the thesis will define and introduce a rigorous scientific method in Internet measurements, (i) by continuously repeating, integrate and correlate data from different Internet “sensors”, (ii) by systematically repeating the analysis that are now done sporadically at best, and (iii) by making access to this data available through the newly defined measurement plane. Special care will be devoted into paying extreme attention to what has been done in close scientific communities, such as the data-base and data-mining domains, to both deal with and make sense out of large volumes of data.
The thesis will be carried out at LINCS http://www.lincs.fr in the context of mPlane, a large Integrated Project (IP) funded by the EU under the 7th Framework Programme, bringing leading experts from academia and industry. Algorithms developed during the thesis may lead to joint implementation with manufacturers (NEC, Alcatel) and experimentation in real operator networks (Orange, Telefonica, Invitel). The methodology that the candidate will follow is, briefly: -* Selection of measurement probes: The candidate will consider some among the software tools early outlined, to be integrated into the measurement plane. Selection of the tools will be performed according to criteria like ease of integration, complementarity of the data provided by the tools, etc. Dressing a thorough list of criteria is an important part of the tasks to be carried on. Notice that the survey and comparison of the different tools, and the identification of the interfaces they can interact through will constitute a research result per se, useful to the whole community. -* Orchestration of measurement probes: The candidate will define and develop a measurement plane, i.e., a software able to collect, analyze data coming from multiple sources, and able to launch new measurements on the basis of the continuously collected results. The candidate will start with few building blocks earlier outlined, and incrementally add complexity and functionalities to the framework. The framework itself should be able to easily accommodate different input/outputs (i.e., tools) as well as new intelligent reasoners (i.e., algorithms). -* Continuous data collection and evaluation: The software should be run on multiple vantage points for a significant portion of the thesis duration, in order to continuously collect, analyze and verify new measurement data. The data will be made available to the scientific community in forms to preserve privacy-related information (e.g., in anonymized or aggregated format). {{References}}


Informations complémentaires (Langue 1)

- mPlane FP7 project, 16 partners in EU. - Cooperation with Jiao Tong University, Shanghai.

Informations complémentaires (Langue 2)

Contact dario.rossi@enst.fr