Proposition de recherche doctorale

Mesure et analyse des services de transport pour les applications cloud

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)

Internet, born as an U.S. military research experiment, is now a worldwide network used by more than three million1 people every day. In addition to interconnecting more and more users, it supports many and diverse applications related to business (e.g., e-commerce, Internet banking), professional activities (e.g., e-mail, videoconferencing, collaborative document editing), education (e.g., e-learning, crowd-sourced encyclopedia), entertainment (e.g., video streaming, on-line gaming), and personal interactions (e.g., instant messaging, social networking, photo sharing). The set of services and platforms related to Internet is continuously expanding, increasing the number of activities that users can conduct online considerably, but also impacting the architecture of the network. Listing few examples: (i) The rise of Content Delivery Networks (1) redefined the paradigm of content distribution and availability; (ii) Cloud Infrastructures (2) are customary for storing and processing huge amount of data; (iii) Mobile Devices (3) are now a common means to access web pages and make use of Internet services; (iv) Cyber Threats (4) are becoming an urgent open issue and a source of troubles, considering the amount of sensitive information being exchanged over the network. Due to the pervasiveness and the heterogeneity of the network, researchers and professionals are facing a constantly renewed interest in understanding the dynamics interacting in such ever-evolving ecosystem. In order to provide a comprehensive understanding, there is the strong need of tools and techniques able to inspect and to characterize various aspects of network traffic. The proposed research activity has its foundations in active measurements campaigns and passive network traffic monitoring. In the former case, active measurements are made possible through the development of ad-hoc test-beds that are instrumented to perform programmable operations and to extract relevant metrics from the network traffic. Operations to be accomplished and meaningful metrics may change according to the scope of the experiment (i.e., different applications types will require different test-beds and measures) and to the final goal of the research. Passive monitoring is instead performed with the advantage of a network probe that observes packets flowing through the monitored link. The probe is required to run a software, e.g. (5), able to reconstruct and abstract traffic flows through the annotation of statistics including (but not limited to) the IP address and the port number used by client and server, the amount of bytes sent and received, the flow duration, the application layer protocol, etc. The network probe needs to be installed in a place where the traffic of a significant amount of hosts (order of thousands) can be monitored. Suitable installation places are, for instance, the Point of Presence of a large-scale ISP, the border router of a University campus, or the link connecting a corporation to the Internet. This is required in order to collect data of a relevant population of users, and prevent unwanted biases caused by the finitude of the collected data. This hybrid methodology allows the study of a huge variety of topics in the computer networks field. The proposed research work focuses on network traffic characterization and measurements applied to three main areas: (i) Cloud Storage Services for personal and data-intensive applications; (ii) Network performance assessment and evaluation of quality perceived by end-users related to the deployment of new protocols and technologies; and (iii) Network events correlation and traffic mining applied to security aspects and end-users protection. In the following a brief description about the intended research on each of the mentioned topics is provided. Cloud Storage In recent years, data storage became a fundamental service with companies, universities and also private end-users having the need of storing large amount. Cloud storage services are emerging as strong alternative to local storage, allowing professional customers to save the costs and the burdens of buying and maintaining expensive hardware (6), (7), and attracting private enthusiasts to backup content with great simplicity and synchronize multiple devices seamlessly (8). Despite the popularity of these services, very little is known about their features: Where are data stored? Which technologies are adopted to transfer them in a secure way? Which is the workload that the cloud infrastructure has to handle? Which traffic load that has to be delivered by the network? Are these services efficient during common usage? Many of these questions still do not find an answer.

Résumé du projet de recherche (Langue 2)

The intention is to propose a methodology to understand and benchmark cloud storage services, unveiling their architecture and capabilities. The usage of an active and configurable test-bed allows the measurement of performance metrics under different workloads, while the understanding of typical usage can be assesses by mean of passive characterization. Network protocols and performance HTTP is the most popular application layer protocol of today's Internet, becoming the de-facto solution for the deployment of new services and applications. However, it is being used to accomplish tasks for which it was not designed for, e.g., to transport video streams, causing inefficiencies in the network (9). Researches are now looking for improvements and solutions in close collaboration with the industry world. Google has recently proposed SPDY, a novel protocol specifically designed to reduce the latency of content transport over the World Wide Web, and proposed it within the Internet Engineering Task Force (IETF) working group HTTP/2 (10). The advantages introduced by SPDY are set back by TCP, which is affected by a high overhead due to connection establishment. To enable to power of SPDY, Google in turn presented a new transport layer protocol named QUIC (Quick UDP Internet Connections) (11), which runs on top of UDP and provides a better support for SPDY. Despite the high interest in these new solutions, a deep analysis evaluating issues and benefits of the SPDY/QUIC is still lacking. Specifically, is there a real advantage in terms of delay when using QUIC in place of TCP? Is this protocol safe to be widely deployed? Does is implement an effective congestion control technique, limiting packet losses? Is it fair with respect to concurrent traffic flows possibly using TCP or plain UDP? Given the novelty and the potential of proposed solutions, several aspects are worth to be investigated. By means of experimental analysis, the effective performance of QUIC will be compared to traditional TCP and UDP protocols in a variety of scenarios, e.g., limited capacity links, interfering traffic, and congested links causing packet losses. Network security Information security over the Internet is becoming a key aspect due to the amount of personal information exchanged every day. Malicious adversaries are increasingly threatening users during ordinary web surfing or mail exchange. Security researchers and practitioners in the field are taking different approaches to detect malware and provide effective countermeasures. For instance, they analyze the instruction set or the malicious code, study the behavior of an infected host in a controlled environment (e.g., a honeypot) (12), or identify malicious connections to Command and Control networks (13). However, these approaches result in a set of methodologies focused on the specific malware being disassembled, but are hard to generalize and make widely applicable. Cyber-attackers modify malware continuously using sophisticated schemes in order to evade detection by security tools, and ,as result, existing antivirus software have a disappointing detection rate (<5%) of newly created viruses (14). Some questions naturally emerge: Is there a way to model malicious activities in general, i.e. without being threat-specific? Is it possible to define and adopt a methodology that addresses the problem from a high-level perspective? How can we distinguish and characterize the benign and the malicious events that might have similar network behavior? Can we automate the process and spot new threats (e.g., zero-day attacks) automatically? The idea is to design a methodology to process, extract and pinpoint network activities taking place with the occurrence of a malicious event. The methodology correlates such activities over time (i.e., analyzing different samples of traffic from the same host) and space (i.e., identifying common patterns among multiple hosts), and leverages information coming from different network layers (e.g., HTTP, DNS, TCP) to uncover hidden malware behavior. The result is an enhanced visibility on the incident that can be framed into a graphical representation, or modeled through the definition of features. Furthermore, such features can be leveraged to distinguish between benign and malicious models, possibly spotting new malicious attacks showing similar properties.

Informations complémentaires (Langue 1)

The thesis will be carried on in an international context, with a co-tutelle agreement among Telecom ParisTech and Politecnico di Torino

Informations complémentaires (Langue 2)

REFERENCES 1. On the use and performance of content distribution networks. Krishnamurthy, Balachander; Willis, Craig; Zhang, Yin;. 2001. ACM Sigcomm Workshop on Internet Measurement. pp. 169-182. 2. A view of cloud computing. Armbrust, Michael; Griffith, Rean; Joseph, Anthony; Katz, Randy; Konwinski, Andy; Lee, Sunho; Patterson, David; Rabkin, Ariel; Stoica, Ion; Zahari, Matei; Fox, Armando. 2010. Communications of the ACM. pp. 50-58. 3. A first look at traffic on smartphone. Falaki, Hossein; Lymberopoulos, Dimitrios; Mahajan, Ratul; Kandula, Strikanth; Estrin, Deborah;. 2010. ACM Internet Measurement Conference. pp. 281-287. 4. A survey on automated dynamic malware-analysis techniques and tools. Egele, Manuel; Scholte, Theodoor; Kirda, Engin; Kruegel, Christopher;. 2012. ACM Computing Surveys. 5. Experiences of Internet traffic monitoring with Tstat. Finamore, Alessandro; Mellia, Marco; Meo, Michela; Munafo', Maurizio; Rossi, Dario. 2011. IEEE Network. pp. 8-14. 6. Cloud storage pricing: a comparison of current practices. Naldi, Maurizio; Mastroeni, Loretta;. 2013. International workshop on Hot Topics in cloud services. pp. 27-34. 7. Amazon S3 for science grids: a viable solution? Palankar, Mayur; Iamnitchi, Adriana; Ripeanu, Matei; Garfinkel, Simson;. 2008. International workshop on Data-aware distributed computing. pp. 55-64. 8. Inside Dropbox: understanding personal cloud storage services. Drago, Idilio; Mellia, Marco; Munafo', Maurizio; Sperotto, Anna; Sadre, Ramin; Pras, Aiko;. 2012. ACM Internet Measurement Conference. pp. 481-494. 9. HTTP as the narrow waist of future internet. Popa, Lucian; Ghodsi, Ali; Stoica, Ion;. 2010. ACM Workshop on Hot Topics in Networks. 10. SPDY Protocol - draft-eitf-httpbis-http2-00. Belshe, Mike; Peon, Roberto; Thomson, Michael; Melnikov, Alexey;. 2012. Internet Engineering Task Force. 11. QUIC, Multiplexed Stream Transport over UDP. Roskind, Jim;. 2012. QUIC: Design Document and Specification Rationale. 12. The Neptnthes platform: an efficiant approach to collect malware. Baecher, Paul; Koetter, Markus; Holz, Thorsten; Dornseif, Maximilian; Frelling, Felix;. 2006. Recent advances in Intrusion detection. pp. 165-184. 13. BotTrack: Tracking Botnets using NetFlow and PageRank. Francois, Jerome; Wang, Shaonan; State, Radu; Engel, Thomas;. 2011. Networking 2011. pp. 1-14. 14. Assessing the effectiveness of antivirus solutions. iMPERVA. 2012. http://www.imperva.com/docs/hii_assessing_the_effectiveness_of_antivirus_solutions.pdf.