4

HIGH PERFORMANCE COMPUTING AND CYBER SECURITY

Mathematical modelling and computer simulation in the fields of ocean, atmosphere, earth science and engineering involve computational tasks which can only be provided by High Performance Computing (HPC). The need for computational power, measured in terms of Giga Floating Point Operations per Seconds (FLOPS), grows exponentially with every bit of increase in the complexity of problem. C-MMACS today has one of the best computing facilities in the country. In addition the HPC group is also involved in research in the field of Cryptography and Network Security.

INSIDE

4.1 Analysis and Characterization of Unsolicited Network Traffic for Denial-of-Service Attack Inference

Source IP address spoofing is the process in which the source IP address of an IP packet is replaced with a 32 bit random number. This technique is widely misused in the Internet for various malicious activities like flooding Denial-of-Service (DoS) attacks. There is an obvious incentive for an attacker in employing source IP spoofing while launching a flood based DoS attack. It enables the attacker to hide its identity, which in turn makes the instant tracking of the attacker to its physical location a tedious process.

Flooding DoS attacks on the Internet with random source IP address spoofing typically triggers a special class of traffic referred as unsolicited packets or Internet background radiations or backscatter. Upon receiving spoofed packets, the victim reacts to such packets and the resulting packets will be spread throughout the Internet address space. By capturing and analyzing these background radiations, significant inference can be made on the nature and prevalence of IP spoofed DoS attacks. Our focus in this work is to identify and characterize background radiations originating from Indian IP address space.

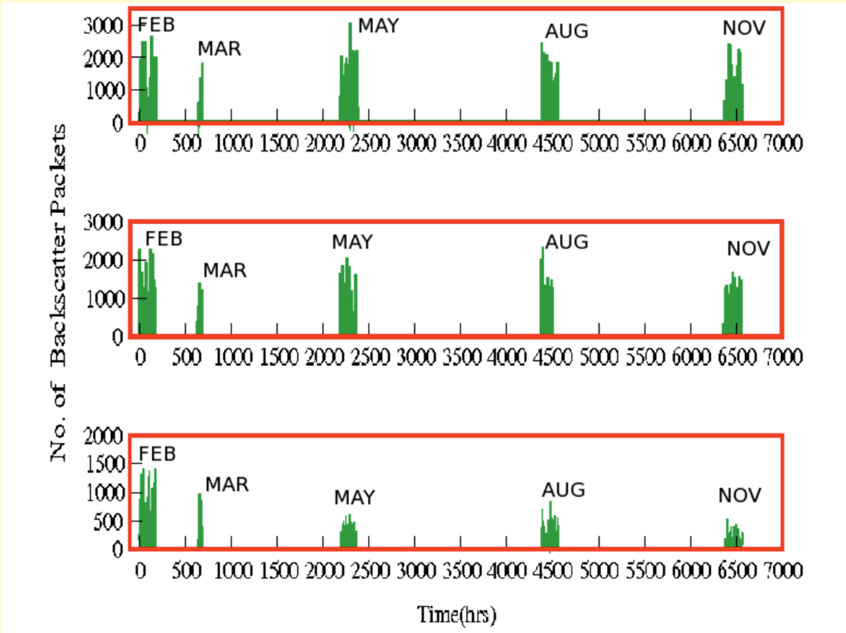


Figure 4.1 Transmission patterns of top 3 malicious flows during year 2008. The blank areas correspond to period for which data was not available. It is important to note that the flow is active in the entire data batch

We started with the global backscatter data collected at Cooperative Associating for Internet Data Analysis (CAIDA) by monitoring 1/256th of the globally routed IP address space. The data consists of quarterly sampled week long backscatter packets for consecutive years from 2005 to 2008. We identified about 3 cores of public IP addresses, which are allocated to Internet Service Providers and other end user organizations in India by Asia Pacific Network Information Centre. A parallel code, based on python and Message Passing Interface, was developed to extract the Indian subset of the global backscatter data. Executing this code on the C-MMACS Altix ICE supercomputer has resulted in a collection of IP-v4 network packets, which are reflections triggered from Indian IP address space as and when they receive source IP spoofed packets. The Indian backscatter subset for year 2008 (37 days of data) is now being analyzed in detail. We have disintegrated these IP-v4 packets and looked into their header part to identify malicious flows targeted to Indian IP space. Preliminary results show that there are significant number of IP spoofed malicious flows just in 37 days in 2008. Figure 4.1 shows the transmission patterns of three top-ranked flows in our list.

V Anil Kumar, V Vinod Kumar, P Sheethal

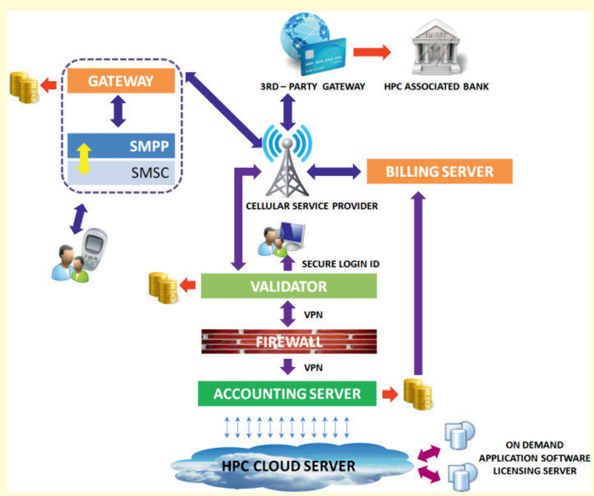
4.2 Framework for Cloud Based Utilities

(a) E-Healthcare through cloud

It has been a real challenge for the government to provide basic healthcare service to the rural areas due to various reasons. One of the main reasons is the unwillingness of expert manpower to go to rural areas due to lack of essential amenities for them as well as their families. The proposed framework for cloud based e-healthcare is an effort to use the latest IT revolution to supplement the existing health services and make them better, faster and cheaper.

Figure 4,2 (a) depicts the proposed framework. Cloud is the latest buzz word in the IT industry, which provides services in the form of pay per use. Here, the cloud provides Healthcare as a Service (HaaS). Eight distinct services are provided by the e-healthcare cloud. Personal Health Records (PHR) is the information that is to be maintained by the patient on the cloud. It is completely protected and can only be accessed by proper authorization from the patient. Electronic Medical Records (EMR) are composed of a clinical data repository that is maintained by the hospital for the patients. This information can be shared by others only with proper authorization from the hospital. Online consultation provides all information about global expert database. Experts availability, consultation cost etc. The other services are namely, E-learning, Medical Alert, Online appointment, Case review and Accounting.

(a)



(b)

Figure 4.2 The proposed framework for (a) e-healthcare (b) Mobile based payment model

The advantage of the proposed framework is that healthcare service can be provided in rural areas which are as good as urban areas. The other advantage is that information about the patient as well as the records can be accessed from anywhere with proper authorizations. This is very advantageous in countries like India, where the population is highly mobile due to the boom in IT centric job opportunities. With already on-going UID projects like “Aadhaar” of the government of India, the e-healthcare implementations are going to be realizable as well as implementable. New features like wellness program, health checkup monitor, drug consultation etc. can be integrated so that it can be made popular among all masses. The emphasis is on utilizing low cost or existing infrastructure through cloud services.

(b) Mobile based payment for HPC Cloud

There has been a growing demand for High Performance Computing (HPC) in non – conventional areas like architecture, graphics design and even gaming over the past few years. Unfortunately, only a few organizations both in government and private enterprise can afford to establish HPC facilities of their own. So the biggest question, in the field of computing is that “Is HPC a dream for a common person's use?” Cloud computing, the future generation computing, could be one solution to realize such dreams. There are many efforts to establish HPC clouds by various institutions. However, many usability constraints have made these HPC clouds unusable for common people. One of the important aspects of any service for common use is the mode of payment. How a common man pays to use a service will decide how popular the service will be. In this work we have proposed, a framework for a mobile based payment model for using HPC cloud. The payment process is very similar to a pre-paid mobile charging, where the user wishing to use HPC goes to his nearest telephone subscriber vendor and purchases a HPC telephone value added card and registers his cellular phone to avail HPC resources. After the registration, the user enables the HPC cellular service by sending the code present in the card to the cellular service provider through a Short Message Service (SMS).

A descriptive view of the model is represented in Figure 4.2 (b). Before the user starts using the HPC, he sends an SMS to the service provider gateway to activate the HPC connections with proper authentication. The service centers provide the Short Message Peer-to-Peer (SMPP) interface and SMS so that billing transactions happen securely via Short message service center (SMSC). The 3rd Party Gateway and the Billing server have a legal agreement with the telephone service provider to access the SMPP interface and the SMS gateway. The SMS gateway is a unique short message service messaging tool which accepts SMS from the users as well as the HPC service provider and stores it in the database as a backup. The SMS gateway, on receiving a SMS from the user immediately passes it on to the cellular service provider billing application. There it checks the user account balance; if it finds sufficient funds to process the request then it makes an entry of the amount owned by the user in the billing server and writes a message in the validator database to allow the particular username.

The user then logs in through the secure login id provided to him from his terminal. On entering the secure login id the validator checks the database for the validity of the user and on successful validation it connects to the firewall through VPN and starts the service. The billing server computes the monetary equivalent of the service usage and informs the cellular service provider to deduct the money from the user account and pay it to the 3rd Party gateway. Once the cellular service provider pays the required amount to the 3rd Party gateway, it immediately alerts the user about the transaction details through an SMS. On receiving the amount from cellular service provider the 3rd Party gateway pays the amount to the HPC Associated bank. The implementation of such an easily accessible payment model will not only enable the common people to use the existing HPC applications but also may find a new market of designing newer and exciting applications using HPC. However, one of the basic implementation challenges is to coordinate among many participating parties and define a viable economic model that will be acceptable and sustainable.

Manmohan Brahma, G K Patra, V Anil Kumar, R P Thangavelu and Rama Krushna Das

4.3 High Performance Computing Resources

In this year CSIR C-MMACS witnessed a major shift in terms of the computing architecture. C-MMACS for the first time procured a high performance cluster to cater to the growing needs of the modelling community of not only C-MMACS but also the whole CSIR.

The HPC cluster is an Altix ICE system with 1152 numbers of processing cores distributed over 96 nodes interconnected in the form of an enhanced hypercube using 4XQDR (40Gbps) infiniband technology. The system, which is currently listed as the 6th fastest (December 2010 list) supercomputer in the country, is equipped with Intel Westmere-EP Hex core processors running at 2.93 GHz frequency with 12 MB L3 cache. Each node has 12 processing cores with 24 GB of memory in a shared memory configuration, while the system as a whole have 2304 GB of memory across the 96 nodes in a distributed architecture. The peak performance the system can achieve is 13.5 TFLOPS, and the sustained performance measured by the High Performance LINPACK (HPL) benchmark is about 11.8 TFLOPS. The system was selected using a strict benchmarking exercise, involving highly computational intensive as well as data intensive models like MoM (Modular Ocean Model) and WRF (Weather Research Forecast). A lustre parallel file system of 30TB was also integrated to the system to handle the storage requirements for the computer systems. Currently this system is being heavily used by National Chemical Laboratories, Pune, Central Electrochemical Research Institute, Karaikudi apart from researchers at C-MMACS and NAL.

Figure 4.3 Altix ICE cluster system with 96 nodes of Intel Westmere-EP Hex core processors running at 2.93 GHz frequency with 12 MB L3 cache. The 6th largest super computer in the country.

The Altix 4700, which continues to be the largest shared memory systems in the country, is being heavily used for specific scientific applications demanding large memory per core. The computing load on these servers are controlled through a workload management software called PBSPro, which not only ensures the efficient usage of the system but also provides a easy user interaction and submission process. The Origin 3900 server continues to provide the home area, while the archival is provided by the high performance SAN (Storage Area Network).

Network Facilities

The C-MMACS HPC facility is utilized by scientists and researchers from C-MMACS and NAL of all the three campus from their desktops through high speed local networks interconnected through a 10Gbps backbone. The researchers from other CSIR laboratories access the facility through an 8 Mbps Internet gateway connected to the Internet through ERNET. C-MMACS continues to provide E-mail facility to 6 nodes from C-MMACS and NAL.

All network servers namely DNS (Domain Name Server), NIS (Network Information Services), WWW (World Wide Web) and Internet gateway have been redesigned and migrated to new servers. These new servers have enabled better services to the C-MMACS users.

C-MMACS is in the process of getting connected to the National Knowledge Network (NKN) with a 1 Gbps Optical Fiber Link. Once all the CSIR laboratories are connected through NKN, it is not only expected to bring better scientific collaborations but also make the C-MMACS HPC facility easily accessible and hence will realize the dream of a centralized HPC facility for CSIR.

High Performance Storage

To support the growing need of data storage, being generated from the data intensive modelling, the existing high performance Storage Area Network (SAN) has been upgraded by procuring an additional Tape library consisting of four numbers of LTO Gen 5 drives. Currently the virtualized 3-tired storage solution has 6 TB online (FC), 20 TB of near-line (SATA) and 360 TB (uncompressed) of offline storage. The SAN provides transparent access to data on all the shared memory HPC systems with hierarchical storage management. This is used as an archival system, while high I/O intensive jobs are run in the local high speed scratch area of the individual servers for better performance.

Software Enhancements

Application software were maintained and upgraded to keep pace with hardware enhancements. The heavily used software are ABAQUS, CFD-ACE+, IDL, GAMIT/GLOBK, Tecplot, S-Plus, Hyperworks, Fluent, ANSYS, OpenFOAM etc. The current list of hardware and software in the computing environment is available at the C-MMACS website http://www.cmmacs.ernet.in. The systems are used extensively for running complex models in the field of ocean, atmosphere and aerospace.

Other Technical Services

Technical support was provided to a large number of users from C-MMACS & NAL. In addition, several students from academic institutions across the country have availed the computing services as part of their academic work at C-MMACS. Technical advice and consultancy were provided to various institutions within and outside CSIR.

R P Thangavelu, G K Patra, V Anilkumar, Ashapurna Marndi and N Prabhu