Threat Modeling Based on Randomized Seeding Attacks In Cloud Virtual Machines

Alexander Ngenzi[1] ,Dr. Selvarani R[2] ,Dr. Suchithra R[3]

[1]*Phd Student , Department Of Computer Science Engineering, Jain University, Bangalore-India*
[2]*Professor ,Computer Science Engineering, Alliance University, Bangalore, India*
[3]*Hod, Department Of Master Of Science In Information Technology, Jain University, Bangalore . India.*

***Abstract**: Threats in virtual machines have been a major challenge in most cloud data centers. The attack/ threat begins from physical machines (hosts) and spreads to all virtual machines(guests). As a result, the virtual machines get infected rapidly by recursive growth of the seeds/ nodes generated in a random manner. This paper proposes threat modeling based on randomized growth of these seeds or nodes. The simulated attacks are free from a deterministic pattern and hence all threats can be detected and prevented. The aim of this work is to develop a mathematical model to prevent seeding attack on virtual machines on the cloud. It presents both Lucas and Fibonacci series and draw relationship between them where by each VM affected is identified and the VMs can be prevented from these attacks.*

***Keywords:** Cloud computing, STRIDE, VMs, APIs, ASF, DREAD, Randomized seeding attack.*

I. Introduction

In this paper, we propose threat modeling " randomized seeding attack Model " that prevents threats/attacks which may affect the virtual machines(VMs) on the cloud. The phrase “cloud” originates from the cloud symbol used by flow charts and diagrams to symbolize the Internet. The term cloud computing refers to both the applications delivered as services over the Internet and the servers and system software in the datacenters that provide those services. The virtual machines(VMs) on the cloud may be affected due to the sharing of resources among themselves. In the proposed model only one virtual machine can be spread all over the other virtual machines and there is a need to analyze the attacks invading the VMs so that we prevent spread of these attacks in the entire cloud. The proposed model is achieved by using both Lucas and Fibonacci series. Both Fibonacci and Lucas numbers are generated by adding the last two numbers in the series. The rest of this paper is organized as follows; it gives literature review/related work of how seeding attack can invade the virtual machines on the cloud , proposes threat modeling system based on random seeding attacks and discusses the experimental setup and results as well as references from different authors on security or threat modeling.

II. Related Work

The customer divides his data among several service providers(SPs) available in the market, based on his available budget. Also it provides a decision for the customer, to which SPs must choose to access data, with respect to data access quality of service offered by the SPs at the location of data retrieval. This does not only rules out the possibility of a SP misusing the customers’ data, breaching the privacy of data, but also can easily ensure the data availability with a better quality of service[1] . In viral marketing, a key problem is to select an initial ”seed” set from the network such that the entire network adopts any behavior given to the seed. Here authors introduced a method for quickly finding seed sets that scales to very large networks. The approach found a set of nodes that guarantees spreading to the entire network under the tipping model[2]. Infrastructure as a Service (IaaS) serves as the foundation layer for the other delivery models, and a lack of security in this layer will certainly affect the other delivery models, i.e., platform as as a service (PaaS), and software as a service (SaaS) that are built upon IaaS layer[7]. From this point of view, the number one service or feature that is missing is security of data. There are two levels of concern here. One is focused on preventing others (such as another customer) from reading private data. This is a clear and obvious concern and prominent in scenarios such as theft, or other direct malicious attack. The other one is concerned with the service provider reading private data. Besides simple lack of trust of the provider themselves, it should be obvious that the service provider is not 100% immune to attacks or other malicious activity, targeted or otherwise. These two levels of concerns apply to other security issues as well, and of course are commensurate with the level of confidentiality desired[8].Authors considered intruder model and requirements that need to be satisfied to provide required level of privacy. Since previous research show that crypto- graphic means cannot always provide protection (especially in long term) authors proposed a trust-based privacy protection. The approach was based on subjective logic that applied to measure/monitor level of trustworthiness of cloud service providers. Authors explained how users have to handle their data to minimize privacy treats in the cloud[9]. In their paper, authors proposed a security metric that enables service providers and service subscribers to quantify the risks that they incur as a result of prevailing security threats and system vulnerabilities[15]. The security metric proposed in this paper was quantified in economic terms, thereby enabling providers and subscribers to weight these risks against rewards, and to assess the cost effectiveness of security countermeasures. Critical to the identification of threats is using a threat categorization methodology. A threat categorization such as spoofing user identity, tempering with data, repudiations, information disclosure, denial of service and elevations of privileges(STRIDE) can be used, or the Application Security Frame (ASF) that defines threat categories such as Auditing & Logging, Authentication, Authorization, Configuration Management, Data Protection in Storage and Transit, Data Validation, Exception Management. The goal of the threat categorization is to help identify threats both from the attacker (STRIDE) and the defensive perspective (ASF). DFDs help to identify the potential threat targets from the attacker's perspective, such as data sources, processes, data flows, and interactions with users. These threats can be identified further as the roots for threat trees; there is one tree for each threat goal. From the defensive perspective, ASF categorization helps to identify the threats as weaknesses of security controls for such threats. Common threat-lists with examples can help in the identification of such threats. Use and abuse cases can illustrate how existing protective measures could be bypassed, or where a lack of such protection exists. The determination of the security risk for each threat can be determined using a value-based risk model such as DREAD or a less subjective qualitative risk model based upon general risk factors (e.g. likelihood and impact)[17]. In this paper, authors used Damage potential, Reproducibility, Exploitability, Affected users and Discoverability (DREAD) modeling.

III. Spreading of randomized seeding Attacks on Cloud

Here we assume that virtual machine VM_1 is the source of attack and carries viruses. Due to the sharing of resources, it affects all other virtual machines and physical machines on the cloud. Once the physical machine is affected, there is a possibility that VM_2, VM_3.........VM_n be affected and process continues until the entire cloud is affected. The orange colored VM is the attack that affects the rest of VMs in the cloud. The green colored VMs are the VMs which have no problem. The affected VM enters in the cloud environment where it causes other VMs to be affected. The outgoing arrows show that the infection is spread all over the VMs on the cloud. This scenario is shown in the figure below taking into account of all VMs:

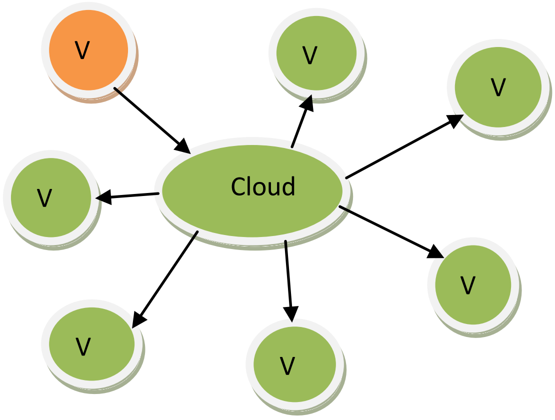

Figure1: Attacks originating from VM 1 are spread over the Cloud

IV. Problem Statement

Threat modeling based on seeding attacks is achieved by using both Lucas and Fibonacci series. Both Fibonacci and Lucas numbers are generated by added the last two numbers in the series. With Fibonacci the series starts with 0,1 where as with Lucas the series starts with 2,1 and so here is part of the Lucas number series:

1. Adding up alternate Fibonacci numbers produces the Lucas numbers. Mathematically

   L(n) = F(n-1) + F(n+1).

   where L is the Lucas series with L(n) is the nth Lucas number. Similarly F for Fibonacci

   2. Adding alternative Lucas numbers produces multiples of the Fibonacci series, where the multiple is 5. i.e.,

   5 F(n) = L(n-1) + L(n+1)

Based on the scenario in the above equations it proves that any individual attack follows Fibonacci series.

V. Proposed threat modeling system

Assume that at least one VM is affected, the attack is seeded onto other VMs. The dynamic sharing of resources among VMs allow each VM to be affected by attacks and eventually affects the entire cloud. In this model, we use Random Fibonacci sequences. This can be demonstrated below:

Given that:

$$L_n = \frac{1}{\gamma}(a_{n-1} + a_{n+1}) \qquad \text{(i)}$$

L_0 =2, L_1 =1

Initially no nodes are affected .$a_0 = 0$

Next " node 1" may be affected.

$a_0 = 0$, $a_1 = 1$ are the first pair of seeds.

L_n is the possible number of nodes to be affected in the nth VM.

The number of nodes affected at the nth machine is following the difference equations given below:

$$a_{n+1} = a_{n-1} + a_n \,,\ n \geq 2 \qquad \text{(ii)}$$

$$a_0 = 0 \,,\ a_1 = 1$$

from (i) above, we assume that All VMs talk to each other and γ is the Random Fibonacci sequence. Equation (ii) is nothing but the Fibonacci sequence.

By using relation between Lucas and Random Fibonacci sequences , we can have the following equations:

1. $\sum_{i=1}^{n} F_i^2 = F_n \ + F_{n+1}$ .... (1)

2. $F_n \ = \frac{1}{\sqrt{5}} \frac{(1+\sqrt{5})^n}{2} + \frac{(1-\sqrt{5})^n}{2}$ .... (2)

3. $L_n = L_{n-1} + L_{n-2},\ n \geq 2$ .... (3)
   $L_0 = 2,\ L_1 = 1$

from the equation (3) above we can get :

$L_n = F_{n-1} + F_{n+1}$ and therefore,

$L_n \ = \frac{(1+\sqrt{5})^n}{2} + \frac{(1-\sqrt{5})^n}{2}$ .... (4)

$\frac{L_j}{L_n}$ is the probability that VM "j" may be affected given that all previous VMs are affected. It is the transmission probability that node j will be affected. Here node and VM are used interchangeably.

Given that:

$$a_n \ = \frac{1}{\sqrt{5}} \left[\frac{(1+\sqrt{5})^n}{2} + \frac{(1-\sqrt{5})^n}{2}\ \right]$$

We need to show how it solves the assumption

$a_{n+1} = \ a_{n-1} + a_n$.

$$a_0 \ = 0\,, \qquad a_1 = 1$$

Proof: $n = 0,\ n = 1$ true (simple substitution by inspection)

Hypothesis: if $n = k$ is true, then $a_{k+1} = a_{k-1} + a_k$

Induction step: if non deterministic sequence(NTS) is true for $n = k + 1$, then $a_{n+2} = a_k + a_{k+1}$

$$RHS = \frac{1}{\sqrt{5}} \left[\frac{(1+\sqrt{5})^k}{2} - \frac{(1-\sqrt{5})^k}{2} + \frac{1}{\sqrt{5}} \left[\frac{(1+\sqrt{5})^{k+1}}{2} - \frac{(1-\sqrt{5})^{k+1}}{2}\ \right]\right.$$

$$= \frac{1}{\sqrt{5}} \left\{\left(\frac{(1+\sqrt{5})^k}{2}\right) \left(1 + \frac{(1+\sqrt{5})}{2}\right) - \left(\frac{(1-\sqrt{5})^k}{2}\right) \left(1 + \frac{(1-\sqrt{5})}{2}\right) \right\} =$$

$$\frac{1}{\sqrt{5}} \left\{\left(\frac{(1+\sqrt{5})^k}{2}\right) \left(\frac{(2+1+\sqrt{5})}{2}\right) - \left(\frac{(1-\sqrt{5})^k}{2}\right) \left(\frac{(2+1-\sqrt{5})}{2}\right) \right\}$$

$$= \frac{1}{\sqrt{5}} \left\{ \frac{(1+\sqrt{5})^k}{2} . \frac{1}{2^2} \left(6 + 2\sqrt{5}\right) - \frac{(1-\sqrt{5})^k}{2} . \frac{1}{2^2} \left(6 - 2\sqrt{5}\right) \right\}$$

$$= \frac{1}{\sqrt{5}} \left\{ \frac{(1+\sqrt{5})^k}{2} . \frac{(1+\sqrt{5})^2}{2} - \frac{(1-\sqrt{5})^k}{2} . \frac{(1-\sqrt{5})^2}{2} \right\}$$

This gives the following equation:

$$a_{k+2} = \frac{1}{\sqrt{5}} \left\{ \frac{(1+\sqrt{5})^{k+2}}{2} - \frac{(1-\sqrt{5})^{k+2}}{2} \right\}$$

for $a_n = \frac{\gamma}{\sqrt{5}} \left\{ \frac{(1+\sqrt{5})^n}{2} - \frac{(1-\sqrt{5})^n}{2} \right\}$

$a_0 = 0,\ a_1 = \gamma$

$$\mathrm{R}a_k \ +a_{k+1} = \frac{\gamma}{\sqrt{5}} \{ \frac{(1+\sqrt{5})^k}{2} \ - \ \frac{(1-\sqrt{5})^k}{2} \ \} \ \ +\frac{\gamma}{\sqrt{5}} \{ \frac{(1+\sqrt{5})^{k+1}}{2} \ - \ \frac{(1-\sqrt{5})^{k+1}}{2} \ \}$$

continue in the same fashion and we obtain

$\gamma \mathrm{a}_{\mathrm{n}+2} = \ \mathrm{a}_{\mathrm{n}} \ = \frac{\gamma}{\sqrt{5}} \{ \frac{(1+\sqrt{5})^n}{2} \ - \ \frac{(1-\sqrt{5})^n}{2} \ \}$

$a_0 = 0, \ a_1 \ = \gamma$

The above equation gives rise to new seeding mechanism

$a_{n+1} \ = a_{n-1} \ + \ a_n$

$a_0 = 0, \ a_1 \ = \gamma \gamma \ \epsilon \, random \ (0{,}1)$

limit Li/Ln = $L_0 - \ 2 \quad (expt \ set \ up)$ i.e. experimental setup

$L_{n+1} = \ \frac{1}{\gamma} L_n \ + L_{n-1} \ \ n \geq 2$ $\qquad\qquad\qquad$ $L_n \ = \gamma \{ \frac{(1+\sqrt{5})^n}{2} \ - \ \frac{(1-\sqrt{5})^n}{2} \ \}$

By comparing with $\mathrm{a}_{\mathrm{n}} \ = \frac{\gamma}{\sqrt{5}} \{ \frac{(1+\sqrt{5})^n}{2} \ - \ \frac{(1-\sqrt{5})^n}{2} \ \}$,

there is relationship between them.

L_n is a solution to the difference equation

$L_{n+1} \quad = \frac{1}{\gamma} (L_n \ + L_{n-1}), \ n \geq 2;$

L_0 =2, L_1 =1 and

$\mathrm{a}_{\mathrm{n}} \ = \frac{\gamma}{\sqrt{5}} \{ \frac{(1+\sqrt{5})^n}{2} \ - \ \frac{(1-\sqrt{5})^n}{2} \ \}$

is a solution to

$a_{n+1} \quad = \frac{1}{\gamma} (a_{n-1} + \ a_{n+1});$

$a_0 = 0, \ a_1 = 1$

$\gamma \ \epsilon \, random \ numbers (0{,}1)$

Assumptions

1. Initially (at 0th instance) no VMs are affected.
2. L_n ≡ Possible number of VMs to be affected.
3. $\frac{L_i}{L_n}$ =Transmission probability.
4. Assume "γ" is the random number between 0 & 1 that a VM request the same resource be attacked

Interpretation: L_0 ,L_1 are the first pair of seeding attack and the consequent attack are modeled by the equation $L_n \quad = \frac{1}{\gamma} (a_{n-1} + \ a_{n+1})$ $\quad L_0$ =2, L_1 =1

Limit= $L_0 - 2 \ (expt \ set \ up)$ i.e experimental setup

As it shown in the figures below, when the value is odd (Figure1),The malicious seeds grow exponentially. The rest of the figures use even number of series attack as it is shown in the figure 1&2.

VI. Experimental setup and discussion of the results

The simulation, set up in Java and GNU Plot is expected to achieve the following goals.

i) The growth of malicious seeds on VMs (L_n versus n)- numerical validation of the theory using randomization.

ii) Vs. n- Observe the transmission/attack probability growth in a dynamic and scalable environment and to ob-serve how "Tail-boosting" helps adjust the sagging probabilities and ensure that the VM's are consistently attacked.

iii) L_n Versus. Time (milliseconds)—to observe how quickly the seeds grow and transmit.

iv) "Tail-boosting" to control the infection and damage caused by the malicious seeds.

v) The growth of the seeds, defined by the Lucas sequence are recursive in nature, hence self-replicating.

vi) The growth of the Lucas seeds is exponential.

vii) In case the Cloud is scaled up which is a very common scenario, the transmission probabilities may go down since the denominator increases while the numerator doesn't. Tail boosting by using an altered transmission.

VII. Result Discussion

The simulation results show the seeding attacks grow exponentially. We present either odd or even number of Lucas sequences which is distinguished by colors in the following graphs. Figure 1 is one of the odd sequences as shown below:

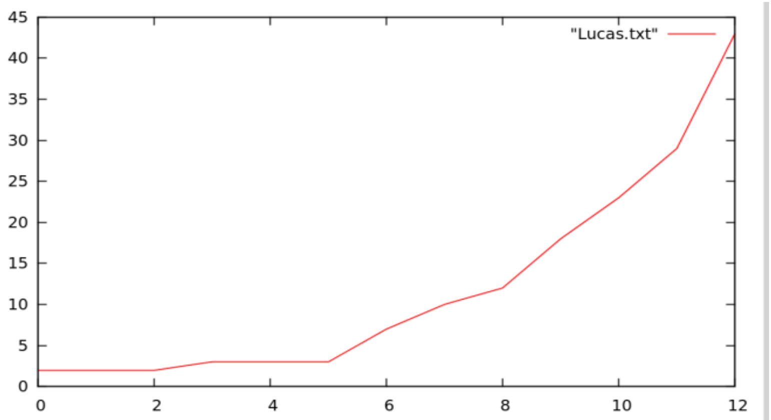

Figure 2: The growth of malicious seeds on VMs (L_n versus *n*) for one seed-set value

Figure 2 below is an example of the even number of the Lucas sequences which is presented by different colors with the respect to the simulated experiment. For example orange, blue, green and gray colors.

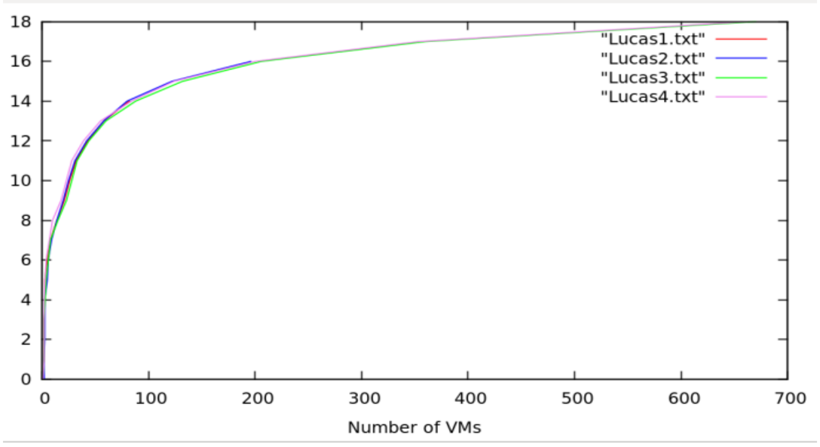

Figure 3: The growth of malicious seeds on VMs (L_n versus n) for n=4, 8, 10, 12

Figure 3 indicates multiple values at n of number of seeds for j number of VMs control of the malicious seeds in a given intervals for example in at j=8 for n= 4.6.8 and 12

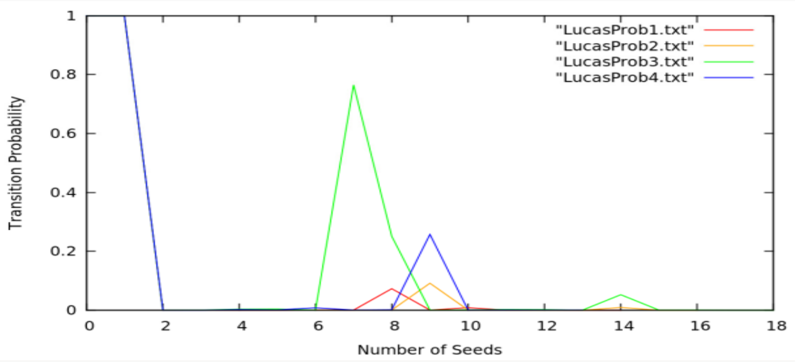

Figure 4: The transition probability on VMs (versus n) for n=4, 8, 10, 12, j=8 and effects of tail-boosting j = Number of dummy VM's for control of malicious seeds.

VIII. Conclusion

The proposed threat modeling system show that the threats/attacks can be prevented from attacking the systems including those of cloud. The simulation and experiments show that the model works perfectly as far as threat modeling is concerned. In this paper, both Lucas and Fibonacci series are used to determine the next threat patterns in the VMs. VM diagram and malicious seed attach graphs are drawn. we discussed how seeding attack can invade the virtual machines on the cloud , propose threat modeling based on random seeding attacks and discussed the experimental setup and results. The future work can be extended to finding how the simulated attacks can be identified in multiple systems on the cloud.

References

[1]. Yu Zhang,Bharat Bhargava (2008) Fibonacci Modeling of Malware Propagation

[2]. Yashaswi Singh, Farah Kandah, Weiyi Zhang (2011); A Secured Cost-effective Multi-Cloud Storage in Cloud Computing. Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conence DOI: 10.1109/INFCOMW.2011.5928887 ,Page(s): 619 - 624

[3]. Wesam Dawoud , Ibrahim Takouna , Christoph Meinel , Infrastructure as a Service Security: Challenges and Solu- tions @2010 in 7th International Conference on Informatics and System, 2010 page 1-8

[4]. Joel Weis and Jim Alves-Foss: Securing Database-as-a-Service: Issues and Compromises, IEEE Security & Privacy,@2011 Volume 9 No-6 page 49-55

[5]. Brent Lagesse," Challenges in Securing the Interface Between the Cloud and Pervasive Systems" IEEE conference, PERCOM @2011 page 106-110.

[6]. Christian Delettre* – Karima Boudaoud – Michel Riveill, Cloud Computing, Security and Data Concealment, Computers and Communication ISCC @2011 IEEE Symposium, page 424-431

[7]. Krešimir Popović, Željko Hocenski (2010), Cloud computing security issues and challenges, @2010, MIPRO 2010 33rd International Covention. Page 344-349

[8]. Vladimir A. Oleshchuk and Geir M. Køien, Security and Privacy in the Cloud A Long-Term View @2011 IEEE.

[9]. Yu Zhang,Bharat Bhargava ,"Fibonacci Modeling of Malware Propagation" IEEE Transaction @2008 , Perdue e-Pubs 08-017.

[10]. Paulo Shakarian · Sean Eyre · Damon Paulo, " A Scalable Heuristic for Viral Marketing Under the Tipping Model" @2013 , arXiv:1309.296

[11]. Z. Chen, L. Gao, and K. Kwiat, Modeling the Spread of Active Worms,Proceedings of the IEEE INFOCOM, 2003

[12]. S. Friedl, Analysis of the new Code Red II Variant, http:// www.unixwiz.net/ techtips/CodeRedII.html, Last ac- cessed Apr 15, 2008

[13]. S. Staniford, V. Paxson and N. Weaver, How to Own the Internet in Your Spare Time, In Proceedings of the 11th USENIX Security Symposium, Aug. 2002

[14]. J. Twycross and M. Williamson: Implementing and Testing a Virus Throttle. In Proceedings of the 12th USENIX Security Symposium, Washington, 2003

[15]. J. Jung, V. Paxson, A.W. Berger, and J. Balakrishnan, Fast Portscan Detection Using Sequential Hypothesis Test- ing, In Proc. of the IEEE Symposium on Security and Privacy, May 2004