Sustainability of Service Provisioning Systems under Attack

Georgios S. Paschos
MIT, USA
CERTH-ITI, Greece
gpasxos@mit.edu

Leandros Tassiulas
University of Thessaly, Greece
leandros@uth.gr

ABSTRACT

We propose a resource allocation model that captures the interaction between legitimate users of a distributed service provisioning system with malicious intruders attempting to disrupt its operation. The system consists of a bank of servers providing service to incoming requests. Malicious intruders generate fake traffic to the servers attempting to degrade service provisioning. Legitimate traffic may be balanced using available mechanisms in order to mitigate the damage from the attack. We characterize the guaranteed region, i.e. the set of legitimate traffic intensities that are sustainable given specific intensities of the fake traffic, under the assumption that the fake traffic is routed using static policies. This assumption will be relaxed, allowing arbitrary routing policies, in the full version of this work.

Categories and Subject Descriptors

H.1 [**Information Systems Applications**]: Models and Principles; Miscellaneous;

General Terms

Algorithms, Reliability, Theory

Keywords

Service provisioning system; guaranteed sustainability; stability

1. SYSTEM MODEL AND DEFINITIONS

Consider a set $\mathcal{N} \triangleq \{1, \dots, N\}$ of parallel servers with constant service rates $\mu_n, n \in \mathcal{N}$. The servers are fed by a set of legitimate streams $\mathcal{L} \triangleq \{1, \dots, |\mathcal{L}|\}$ of *traffic*, each stream $l \in \mathcal{L}$ associated with traffic intensity a_l and a set of reachable servers $\mathcal{S}_l \subseteq \mathcal{N}$. The traffic arriving from a stream l is *routed* to some of the servers in $\mathcal{S}_l$.

A malicious system launches a Degradation of Service attack (a type of Denial of Service attack) in order to disrupt the operation of the system. In particular, the malicious system has a set $\mathcal{M} \triangleq \{1, \dots, |\mathcal{M}|\}$ of malicious traffic streams, where the stream $m \in \mathcal{M}$ generates fake traffic with intensity b_m and is capable of routing it towards a subset of servers $\mathcal{Q}_m \subseteq \mathcal{N}$. See Figure 1 for an example of the studied system in terms of a bipartite graph.

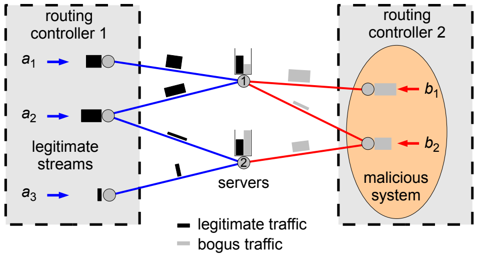

Figure 1: An example of the system for 2 servers, 3 legitimate streams and 2 malicious streams. Also, $\mathcal{S}_1 = \mathcal{Q}_1 = \{1\}$, $\mathcal{S}_2 = \mathcal{Q}_2 = \{1, 2\}$ and $\mathcal{S}_3 = \{2\}$.

We assume the operation of two controllers with conflicting interests. Controller 1 splits legitimate traffic to allowable servers according to *routing coefficients* $f_{ln}, (l, n) \in \mathcal{L} \times \mathcal{N}$. We collect all policies that satisfy $\sum_{n \in \mathcal{N}} f_{ln} = a_l$ and $f_{ln} = 0$, if $n \notin \mathcal{S}_l$ in the *feasible* set Π_1. Controller 2 operates in a similar manner, choosing coefficients ϕ_{mn}, $(m, n) \in \mathcal{M} \times \mathcal{N}$ to satisfy $\sum_{n \in \mathcal{N}} \phi_{mn} = b_m$ and $\phi_{mn} = 0$, if $n \notin \mathcal{Q}_m$ for all m. Π_2 is the set of malicious policies.

The typical stability condition for a server reads: a server n is stable iff the aggregate arrival intensity is smaller or equal to its service rate; this is referred to as *rate stability*. From the practical viewpoint, though, the DEGoS attack is considered successful only if service to legitimate traffic fails. If some servers are unstable in the traditional sense but they are avoided by the legitimate traffic then the attack has failed to harm the system. Thus, we slightly change the definition of stability as follows:

DEFINITION 1. (System Stability) *A server $n \in \mathcal{N}$ is stable if*

$$\sum_{l \in \mathcal{L}} f_{ln} + \sum_{m \in \mathcal{M}} \phi_{mn} \leq \mu_n$$

or if $\sum_{l \in \mathcal{L}} f_{ln} = 0$. The system is stable if all servers are stable.

Let $\mathbf{a} \triangleq (a_1, \dots, a_{|\mathcal{L}|})$ denote the vector of legitimate traffic intensities. We extend the standard notion of system stability region to include the impact of a malicious intruder with fake traffic intensities $\mathbf{b} \triangleq (b_1, \dots, b_{|\mathcal{M}|})$ and policy $\boldsymbol{\phi}$.

DEFINITION 2. (Sustainable region $\Lambda_{\mathbf{b}}^{\boldsymbol{\phi}}$) *The sustainable region $\Lambda_{\mathbf{b}}^{\boldsymbol{\phi}}$, when the malicious adversary operates with a malicious policy $\boldsymbol{\phi} \in \Pi_2$ and available fake traffic intensities $\mathbf{b}$, is the set of all $\mathbf{a}$ for which there exists a legitimate policy $\mathbf{f} \in \Pi_1$ such that the system is stable.*

Moreover, we define the notion of guaranteed sustainable (or simply "guaranteed") region as the set of legitimate traffic intensities $\mathbf{a}$ which are guaranteed to be sustainable regardless of the malicious policy used.

SIGMETRICS'13, June 17–21, 2013, Pittsburgh, PA, USA.
ACM 978-1-4503-1900-3/13/06.

Definition 3. (Guaranteed region $\Lambda_{\mathbf{b}}$) *The guaranteed region $\Lambda_{\mathbf{b}}$ of the system attacked by a malicious adversary with available traffic intensities* $\mathbf{b}$*, is the set of all* $\mathbf{a}$ *for which there exists a legitimate policy* $\mathbf{f} \in \Pi_1$ *such that the system remains stable* ***under any selection*** $\boldsymbol{\phi} \in \Pi_2$.

The guaranteed region is parametrized by the fake traffic intensity, $\mathbf{b}$. For $\mathbf{b}$ large enough, $\Lambda_{\mathbf{b}}$ might contain only the zero element vector $\mathbf{0} \triangleq (0, 0, \ldots, 0)$, which implies that there is a malicious policy $\boldsymbol{\phi}$ such that even arbitrarily small legitimate traffic intensities are not sustainable, regardless of the legitimate policy $\mathbf{f}$ used. In practical terms, we can think of such a situation as a DoS attack. *The DEGoS attack, on the other hand, corresponds to cases where the guaranteed region is not degenerated and legitimate traffic can still be sustained despite the attack, albeit in smaller intensities.*

2. MAIN RESULT

First, we fix a malicious policy $\boldsymbol{\phi}$ and study the sustainable region of traffic intensities under this policy. Let $r_n(\boldsymbol{\phi}) \triangleq \left(\mu_n - \sum_{m \in \mathcal{M}} \phi_{mn}\right)^+$ be the *available resource* of server n after the traffic arriving from malicious streams under $\boldsymbol{\phi}$ is subtracted. We use $(.)^+ \triangleq \max\{., 0\}$. Using the stability definition, we conclude that the system is stable iff there exists a legitimate policy $\mathbf{f}$ such that

$$\sum_{l \in \mathcal{L}} f_{ln} \leq r_n(\boldsymbol{\phi}), \text{ for all } n \in \mathcal{N}. \qquad (1)$$

In what follows, we will express the sustainable region $\Lambda_{\mathbf{b}}^{\boldsymbol{\phi}}$ in terms of traffic intensities $\mathbf{a}, \mathbf{b}$ and service rates $\boldsymbol{\mu}$. For an arbitrary non-empty subset of the servers $\hat{\mathcal{N}} \subseteq \mathcal{N}$ consider the induced subsets $\hat{\mathcal{L}}, \hat{\mathcal{M}}$, where

- $\hat{\mathcal{L}} = \left\{ l \in \mathcal{L} : \mathcal{S}_l \subseteq \hat{\mathcal{N}} \right\}$ is the set of legitimate traffic streams that **must** direct all traffic to some of the servers in $\hat{\mathcal{N}}$ and
- $\hat{\mathcal{M}} = \left\{ m \in \mathcal{M} : \mathcal{Q}_m \cap \hat{\mathcal{N}} \neq \emptyset \right\}$ is the set of fake traffic streams that **can** direct fake traffic to some of the servers in $\hat{\mathcal{N}}$.

Lemma 1 (Cut constraints). *The traffic intensities* $\mathbf{a}$ *are sustainable under* $\boldsymbol{\phi}$ *if and only if*

$$\sum_{l \in \hat{\mathcal{L}}} a_l \leq \sum_{n \in \hat{\mathcal{N}}} r_n(\boldsymbol{\phi}), \qquad \textit{for all } \hat{\mathcal{N}} \subseteq \mathcal{N}.$$

2.1 Guaranteed region $\Lambda_{\mathbf{b}}$

Consider an auxiliary network $\mathcal{G}(\hat{\mathcal{N}}) = (\mathcal{V}, \mathcal{E})$. We define the set of nodes as $\mathcal{V} \triangleq \{s, t, u_i, v_j : i \in \hat{\mathcal{M}}, j \in \hat{\mathcal{N}}\}$, where s is the source node, t is the sink, nodes $u_i, i \in \hat{\mathcal{M}}$ correspond to members of $\hat{\mathcal{M}}$ and nodes $v_j, j \in \hat{\mathcal{N}}$ correspond to members of $\hat{\mathcal{N}}$. The set of links consists of three subsets $\mathcal{E} = \mathcal{E}_\mu \cup \mathcal{E}_Q \cup \mathcal{E}_b$, where each subset consists of directional links defined as follows

$$\mathcal{E}_b \triangleq \{(s, u_i) : i \in \hat{\mathcal{M}}\}, \mathcal{E}_\mu \triangleq \{(v_j, t) : j \in \hat{\mathcal{N}}\}$$
$$\mathcal{E}_Q \triangleq \{(u_i, v_j) : i \in \hat{\mathcal{M}},\ j \in \mathcal{Q}_i\}.$$

A link (s, u_i) has capacity b_i, a link (v_i, t) has capacity μ_i, while all links in subset $\mathcal{E}_Q$ have infinite capacity. Let $M_{\max}(\hat{\mathcal{N}})$ denote the maximum s-t flow of network $\mathcal{G}(\hat{\mathcal{N}})$.

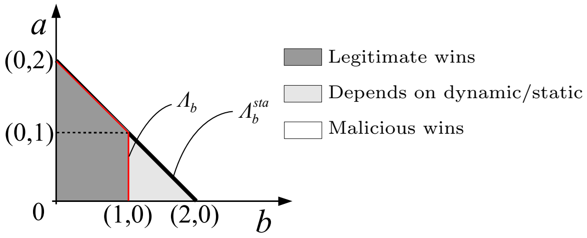

Figure 2: Regions of the studied example for the case of static policies (Λ_b^{sta}) and for the dynamic (Λ_b). The sensitivity of the guaranteed sustainability to dynamic malicious policies is visible.

Definition 4 (Conditions C.1). *The following inequality is satisfied*

$$\sum_{l \in \hat{\mathcal{L}}} a_l \leq \sum_{n \in \hat{\mathcal{N}}} \mu_n - M_{max}(\hat{\mathcal{N}}), \qquad \textit{for all } \hat{\mathcal{N}} \subseteq \mathcal{N}. \qquad (2)$$

Theorem 1. (Guaranteed region) *Conditions C.1 are necessary and sufficient to guarantee sustainability for the traffic intensity* $\mathbf{a}$ *under any* $\boldsymbol{\phi}$*.*

3. DISCUSSION OF THE DYNAMIC CASE

In the followup work we extend the study to the case of dynamic routing polices. In case controller 2 is static, *Join the Shortest Queue* turns out to be the optimal policy for the legitimate controller and it can be shown that the guaranteed region described here is achieved by this policy. If, however, the controller 2 is allowed to allocate bogus jobs in a dynamic fashion, the guaranteed region changes drastically. Below we demonstrate this in an example.

3.1 An example with two servers

Consider two servers with unit service rate fed by one legitimate stream with traffic a and one malicious with traffic b. Traffic can be routed to both servers. Using the results of the previous section, we conclude that $a + b \leq 2$ is a necessary and sufficient condition for guaranteed sustainability as long as the malicious intruder is constrained to static routing policies. We call this region Λ_b^{sta}, see Figure 2.

Definition 5 (Switching malicious policies). *A switching malicious policy directs all fake traffic to one server during a time interval of length τ_i, alternating the server in each interval. During the ith interval, $i = 1, 2, \ldots,$ the fake traffic is directed to server $1 + (i + 1 \mod 2)$. The duration of the ith interval is given by the sequence τ_i, $i = 1, 2, \ldots$.*

Theorem 2 (Region under dynamic policies). *The guaranteed region for the example of two unit servers is*

$$\begin{array}{ll} a + b \leq 2 & \textit{if } b \leq 1 \\ a = 0 & \textit{if } b > 1. \end{array}$$

Examples of switching malicious policies that intuitively lead to the above result are: $\tau_i = i$ and $\tau_i = 2^i$.

4. ACKNOWLEDGMENTS

The work of G. Paschos is supported by the WiNC project of the Action:Supporting Postdoctoral Researchers, funded by national and Community funds (European Social Fund).