Position Paper Title: Denial of Information Attacks

Calton Pu

College of Computing, Georgia Institute of Technology

1. Introduction

Quality of Service (QoS) is an important requirement in many high performance transaction systems. An example is "95% of transactions must complete within 2 seconds of submission". Another example is "the system availability should achieve five 9's". In large distributed systems, the term QoS also has included end-to-end performance metrics (e.g., guaranteed network bandwidth and latency). Recently, Denial of Service (DOS) attacks have become a major threat to system QoS. DOS attacks aim to reduce the QoS available to legitimate users by saturating some system resource through a flood of syntactically correct requests.

Many Internet and HPTS applications, e.g., digital libraries and electronic commerce, are built around information flows. Their main goal is to transport the right information to the right user at the right time. From school children to experts who manage critical national scale systems, an increasing number of information consumers are depending on information content that is relevant, accurate and satisfactory in serving the request. We believe that providing *Quality of Information* (QoI) in large distributed information flow applications is a requirement that will follow the QoS requirements. In analogy to the many dimensions of QoS, there are also many dimensions of QoI, such as the consistency, timeliness, reliability, trustworthiness, and density/richness of information.

A fundamental assumption made by many information rich applications is the ability of the database system to find and deliver information with satisfactory QoI when such information is needed. This assumption is vulnerable to the intentional introduction of noise in the information system to confuse (or lower the efficiency of) the mechanisms for finding resources and information. We call this noise introduction *denial of information* (DOI) attack, which is the information analog of DOS attacks. Similar to DOS, which floods a particular service with massive syntactically correct requests, DOI floods resource discovery and information services by diluting their content with massive noise data that is syntactically correct. (See examples in Section 2.) Consequently, even when critical information is accessible in principle, it may be difficult or impossible to find it, resulting in lower or unacceptable level of QoI.

Although DOI attacks are the information analog of DOS attacks, there are some important differences between them. For example, DOS attacks must be massive attacks that overwhelm the server capacity, thus degrading the service level. In contrast, DOI attacks may be either massive or gradual. A massive DOI attack generates a large amount of spurious information, hoping that the information system will transmit a portion of it to the consumer and hamper the consumer's ability to discern the actual picture of reality, at least momentarily. A massive DOI attack is usually unconcerned with detection and is designed to produce immediate confusion before recovery can take place. In contrast, a gradual DOI attack attempts to remain below the detection threshold and continues to disseminate small amounts of spurious information surreptitiously. Over a period of time, a gradual DOI attack may be able to trick the information system to deliver increasingly misleading information. Finally, even if a system has

a perfect defense against DOS attacks, it is still vulnerable to both gradual and massive DOI attacks.

We believe that QoI assurance is an important security requirement that needs to be addressed by high performance information systems in the near future. Traditional information security deals with confidentiality, integrity and availability of information, but they only capture a subset of QoI dimensions. For example, integrity typically means that information is not altered by unauthorized parties. Integrity captures QoI only if information is accessed from known and authenticated sources. When applications must discover relevant information, the mere accessibility of critical information becomes insufficient if the user must find the needle in a haystack of noise introduced by DOI attacks that effectively hide the needed information. In the long term, DOI attacks may be more damaging than DOS because they can succeed even with a perfect defense against DOS (see Section 3).

2. Examples and Models of DOI Attacks

We observe that computer threats tend to be under-appreciated during an initial incubation period, characterized by an ongoing debate on whether those threats are "real". For example, although security experts knew about many Unix exploits for several years, it was the Morris Worm that made network security an issue with the public and convinced many company management teams to invest in security. Similarly, technicians understood the importance of distributed denial of service (DDOS) threats after the wave of SYN attacks in mid-90's. However, the public awareness and management recognition of DDOS threats came only after the successful DDOS attacks that brought down high profile services such as Yahoo!, CNN, and others. What we observe is an incubation period between the introduction of a threat and a catastrophic failure of the network when a capable hacker finds an efficient way to propagate the "mature" threat. Given the past trend, it is safe to say that after a threat becomes known, eventually it will be used in a large scale attack. It is a matter of when, not if.

In terms of DOI attacks, we are still in the incubation period. We have recognized the threat, but it has not been used in a large scale attack. However, there are many indications that the threat is real. In this section, we discuss some model-examples of DOI attacks, where noise is intentionally introduced into the database to lower the quality of information. We use the term "model" in the biological science sense, where a "mouse model" means using the treatment of a disease in mice to show potential effectiveness of such treatment in analogous diseases in humans. Although we have yet to see a large scale DOI attack on a mission-critical database, we will discuss examples and models of inserting noise and misinformation into representative information systems.

Consider the use of grocery chain cards (Safeway and Kroger are well known examples) as a data warehouse quality assurance tool. In exchange for high quality information about a family's shopping profile and to encourage a frequent shopper, the grocery chain is willing to give significant financial incentives in the form of discounts. While they enjoy the discounts made available, some people are concerns about their privacy due to the time scale and level of details in the data warehouse. In response, several strategies have been devised to counteract this information gathering. For example, some shoppers simply request a new card every few months with a different spelling of names or addresses. The loss of a card limits the length of time shopping information can be aggregated for analytical processing. Another strategy is the reported "lottery" exchange of cards, where a large group of people (e.g., in a conference) throw

their cards into a basket, and then pick up randomly a card from the basket. This strategy further lowers the quality of information for that group of card holders, since the buying habits of unrelated owners are now joined together. Once this kind of noise is introduced into the data warehouse, it is virtually impossible to cleanse the information.

Another example of information scrambling is a typical insider attack. A disgruntled employee, for example, might delete computer files or records from a database (or physically shred paper files). An attacker knowledgeable about backups, knowing the futility of simply deleting files, may insert random records into a database that are harder to find and clean up. A more sophisticated attacker may change some fields in random records, since the smaller a change is, the harder it is to trace and repair. Some reports put insider attack damages at an average of millions of dollars. The high cost of such attacks may be partially due to the difficulties in finding and repairing the damage.

An early example of massive DOI attacks is the so-called "mail bomb". On August 10, 1996, Mr. Dave Methvin, an executive editor of Windows magazine at that time, was one of more than a dozen persons who suffered a typical mail bomb attack. The list of persons attacked included former President Bill Clinton and Bill Gates. The attacker subscribed the victims to over 1000 mail lists. Mr. Methvin is a good example of how a massive DOI attack can impair someone from getting information, in this case email. Just 10 hours after the attack started, he was flooded with 1600 mail messages. Although most of mailing lists now require confirmation of subscription requests to prevent this specific problem, there are many variations of this attack that can seriously affect a user's capability to read email in critical situations.

To detect such massive DOI attacks, we need to define appropriate QoI metrics that can help us distinguish between normal processing and under-attack situations. One approach is to adapt and apply techniques from intrusion detection, where system QoS parameters are monitored to distinguish normal behavior from deviations that indicate an ongoing attack. The basic premise for anomaly detection is that there is intrinsic and observable characteristic (or regularity) of normal behavior that is distinct from that of abnormal behavior. By defining QoI metrics and the information system's normal behavior, we can establish thresholds beyond which a DOI attack may be under way. For example, a user who receives about 100 email messages daily, the arrival of 200 emails within 2 hours may indicate a mailbomb attack.

Related to mailbomb, another example of intentionally introducing noise into an information base is *spam*, which has been growing exponentially in recent months. While emails are usually stored in files, not databases, email messages constitute an important source of information for all of us. As reported in the press recently, large ISPs such as AOL and Hotmail have seen the percentage of spam messages exceed 50% and climbing. With a total number of emails processed at more than a billion a day, the resources spent on processing and storing spam is staggering. However, the main problem for the recipients is not the processing and storage of spam, but the human attention required to filter out the spam. For those of us who receive more than a hundred emails a day, it is a significant amount of effort to delete the spam messages. Although spam is not a catastrophic DOI attack in our definition, it is a useful model through which we can learn more about the effectiveness of some defense strategies.

We can classify the defenses against spam into three groups. The first group consists of content-based filters. These tools are similar to virus detectors, since they use signatures to identify spam messages and filter them out before they reach a user's inbox. The main difficulty with filters is that spam creators can easily change parts of the spam message to bypass any

specific filter. This is particularly evident in the spam messages that consist almost entirely of images, which are virtually impossible to filter out. The second group consists of explicit management of black lists (reject messages from known bad senders) and white lists (only accept messages from known good senders). Black lists are considered ineffective due to address spoofing. White lists consider everyone to be guilty until proven innocent, which makes reconfiguration and adaptation very cumbersome, among other problems. The third group introduces some costs for email exchanges to make spam (a form of junk mail) more expensive. We will discuss this approach in the next section.

To counter DOI attacks such as spam, an alternative approach (a generalization of black/white lists) is to measure and maintain the reliability, authority, or trustworthiness of information sources as a QoI dimension. This is an important problem that has received considerable attention in the Web community. Zagat Survey, for example, is considered an authoritative source on the ranking of restaurants due to the variety of information sources included in the survey. Other popular web sites such as CNET use similar methods to establish their credibility. How to maintain and propagate such trust measure with the information is an open research area. For example, if an e-commerce application purchases certain goods on the web based on information made available by vendors, it could use a trust measure that reflects the ability of the online vendor to supply the merchandise of agreed upon quality at the advertised price in a timely manner. In another setting, if a scientist is searching for information relevant to an experiment that she is conducting, the trust measure may reflect the belief that the data found in a scientific publication has been peer-reviewed and thus is scientifically valid. The goal of the trust measure is to filter out noise and misinformation by associating low values of trust with unknown sources, while maintaining high levels of trust with information objects produced by reliable sources.

3. The High Performance Paradox

For DOS attacks, faster computers have been part of the defense. The faster and bigger a machine is, the harder it is to slow it down through flooding of requests. Usually, the added capacity would give the defense more time to react to a DOS attack. Therefore, over-provisioning can be seen as a brute force approach to DOS defense. Ironically, the faster a system can process new information, the more damage it can do under a DOI attack. We call this phenomenon the *High Performance Paradox*.

As a concrete example, consider two transaction processing systems, system A running at 100 transactions per second (TPS) and system B at 500 TPS. Suppose that someone starts a DOI attack on both systems A and B simultaneously, sending a bogus transaction stream to each system and succeeding in using up 50% of the workload (50 TPS for system A and 250 TPS for system B) for the DOI attack stream. Let us assume that the DOI attack is discovered 10 minutes after it started for both machines, so the attack lasts only 10 minutes on each machine. (This scenario is optimistic for current state-of-art in system survivability and defense against DOS/DOI attacks. But the important point here is that the attacks are discovered at the same time for a fair comparison, not the absolute timing.) Let us further assume that on average, each DOI attack transaction infects 5 data items with wrong information. Consequently, system A would have processed 15,000 bogus data items, while system B would have processed 75,000. Since the recovery from information contamination is currently dominated by manual repair, it is

easy to see that the faster system (B) would have been affected for longer time than a slower system.

Due to the High Performance Paradox, the core component of many DOI defenses is an ability to slow down the execution of jobs in the system. In the case of spam, the goal of the defense is to slow down the attacker (sender of spam). This is the basis for the third group of defense against spam mentioned above. One way to achieve the slow down is to give senders explicit tasks to perform (e.g., a puzzle that consumes significant amount of CPU for each email delivery in the Pennyblack project at Microsoft Research), therefore making the spammers less effective.

Traditionally, high performance information processing systems were only concerned about running faster and faster. When considering threats such as DOI, it becomes apparent that high performance requires an ability to run both faster and slower, selectively.

4. Summary

In this position paper, we outline the Denial of Information (DOI) problem, the information analog of denial of service (DOS). We summarize some examples of DOI attacks, including the mailbomb and spam. Although there have not been high profile and massive losses due to DOI attacks, we believe it is a serious threat for high performance information systems. On the problem side, there are significant similarities and differences between DOS and DOI attacks. The similarities create opportunities for research that adapt DOS defense techniques for DOI. The differences (e.g., the High Performance Paradox described in Section 3) create opportunities for research that go beyond DOS.

On the solution side, we briefly summarize some approaches that can help defend against DOI attacks. The first one is the adaptation of intrusion detection techniques based on information theoretic approaches to detect massive DOI attacks such as the mailbomb. The second one is the development of trust management techniques to determine the level of reliability and trustworthiness of information sources to counter DOI attacks such as spam. In addition, the High Performance Paradox shows that we need the ability to make a high performance information system run both faster and slower, as determined by the situation.

Acknowledgement. The Georgia Tech team investigating the Denial of Information problem is supported by an NSF ITR grant (CCR-0121643). The senior personnel consist of Calton Pu, Mustaque Ahamad, Wenke Lee, Kang Li, Ling Liu, Leo Mark, and Edward Omiecinski. Graduate students that have been working on the project include W. Huang, X. Li, and L. Singaravelu.