ABSTRACT

This paper introduces FOREVER, a novel service that can be used to enhance the resilience of replicated systems, namely those exposed to malicious attacks. The main objective of FOREVER is to remove faults and intrusions that may happen during system execution, and such removal is done by combining both evolution and recovery techniques. The paper presents (i.) the challenges that systems exposed to malicious attacks need to address, and (ii.) how FOREVER can be used to tackle these challenges.

CATEGORIES AND SUBJECT DESCRIPTORS


GENERAL TERMS

Security, Reliability

KEYWORDS

Intrusion tolerance, rejuvenation, diversity

1. INTRODUCTION

Attackers are actively involved in the development of new techniques to inject or activate latent faults in existing systems [1], which implies that threats evolve during the lifetime of systems exposed to malicious attacks. This means that such systems need also to evolve in order that attacks never originate system failures. The ideal goal of such evolution would be the complete removal of vulnerabilities (thus eliminating any chances of an attack causing a failure), but it is well known that such goal is very difficult, if not impossible, to achieve. Nevertheless, during system execution, one can minimize the number of vulnerabilities by applying security patches to operating systems or by introducing newer (better) versions of application code.

The remaining vulnerabilities may be targeted by attacks and produce faults or intrusions. A resilient system needs to deal with such faults/intrusions, which may be masked through Byzantine fault-tolerant (BFT) replication protocols (e.g., [6, 18, 16]). These protocols are typically run on replicated systems and are able to tolerate the failure of a finite set of f replicas. The problem is that given a sufficient amount of time, a malicious adversary can find ways to compromise more than f replicas. Therefore, if one wants to build a resilient system that is continuously operating, then some sort of recovery mechanism will need to be added. The goal would be to detect and (reactively) recover compromised replicas at a pace faster than the time needed by an adversary to compromise more than f replicas. However, arbitrary faults are very difficult to detect [7, 11]. One alternative approach is to estimate the minimum time needed by an adversary to compromise more than f replicas and (proactively) trigger periodic replica recoveries at a faster pace [20, 30, 6]. Note that the time needed by an adversary to compromise more than f replicas is highly dependent on how different/diverse replicas are [19], and the use of diversity in security raises several concerns while a large amount of work remains to be done in order to get a workable solution [1].

The simplest recovery procedure is to boot a clean image containing the original operating system and application(s) code, and to obtain the current state (if any) from the surviving replicas. Clearly this technique removes the effects of any faults/intrusions that could have occurred before the recovery. However, the bugs/vulnerabilities that caused such faults/intrusions may remain or the adversary may have acquired knowledge before the recovery (e.g., user passwords, OS version) sufficient to deploy a more advanced attack after the recovery. Following this reasoning, the adversary may accumulate knowledge over time (days, weeks, months) until he is able to compromise more than f replicas between recoveries. This means that diversity in the space domain should be complemented with diversity in the time domain: recoveries should introduce diversity.

This paper presents the Fault/intrusiOn REMoVal through Evolution & Recovery (FOREVER) service, which addresses some of the challenges identified above. This service can be used to enhance the resilience of replicated systems, namely those that can be affected by malicious attacks. As the name implies, the main objective of FOREVER is to re-
move faults and intrusions by combining both evolution and recovery techniques.

2. CHALLENGES

In this section we analyze the main challenges that any system exposed to malicious attacks needs to address.

2.1 Threats Evolve During System Lifetime

If a system needs to be deployed in an environment (e.g., Internet) where malicious attackers may be present, then it is certainly not safe to assume that threats will be always the same. In fact, it is well-known that attackers are actively involved in the development of new techniques to inject or activate latent faults in existing systems [1].

The usual way of fighting such malicious behavior is by applying security patches to operating systems or by introducing newer (better) versions of the application code. Typically, these activities are done manually by a system administrator and may introduce unavailability periods in system operation.

We argue that money-critical (e.g., online banking, e-commerce websites) and safety-critical (e.g., critical infrastructures) systems should not depend on human intervention and that unavailability should be avoided at all costs. In fact, our main concern is with safety-critical systems, given the recent shift in attackers’ attention from classical money-critical targets, such as e-commerce websites, to safety-critical infrastructures (e.g., power, water, gas) [5, 17].

2.2 BFT Protocols are Not Enough

Byzantine fault-tolerant protocols may be used in replicated systems to deal with the arbitrary failure of a finite set of \( f \) replicas [6, 18, 16]. Such protocols have limited utility in long-lived systems where malicious adversaries are constantly deploying attacks and causing intrusions. In fact, fault/intrusion tolerant protocols are useful to delay the corruption of the overall replicated system by a certain interval of time that depends on the actual value of \( f \) and on how different/diverse the individual replicas are [19]. Unfortunately, while a higher \( f \) allows to tolerate more compromised replicas, it also means that there are more individual replicas that need to be different from each other. For instance, most Byzantine fault-tolerant consensus algorithms (a fundamental building block for replication protocols) for practical environments require a minimum of \( 3f + 1 \) processes to tolerate faults on at most \( f \) of these. This means that when \( f \) is increased from 1 to 2, the minimum total number of replicas is increased from 4 to 7. Notice that if these three extra replicas are not different from the existing ones, we are in fact increasing the probability of the overall system becoming compromised.

We argue that fault/intrusion tolerant protocols are useful as long as they are complemented with recovery mechanisms able to reduce both the value of \( f \) and the time window for an adversary to compromise more than \( f \) replicas.

2.3 Recoveries may be Attacked

As explained in the previous section, recovery mechanisms may play an important role in ensuring that no more than a certain number of replicas is compromised at the same time. Some existing works implement proactive recovery making the system tolerate an unbounded number of malicious faults during its lifetime [6, 30].

If a system has these mechanisms deployed, they will be the natural first target of any attack: a malicious adversary will start by stopping, delaying, or even corrupting the recovery process, before deploying the actual attack that compromises the replicated system. A recent work shows how a set of existing systems may collapse precisely because the recovery mechanisms are not adequately protected [26]. This work analyzes systems that are designed under the asynchronous model, but end-up making a few synchrony assumptions, namely related to recoveries triggering and execution time. Such assumptions may have acceptable coverage in the presence of accidental (non-intentional) faults, but represent a vulnerability in a malicious environment.

We argue that recovery mechanisms must be architectured in a way that their timeliness and effectiveness cannot be affected, namely by a malicious adversary.

2.4 Recoveries Affect Availability

When a replica is recovering, there exists normally a non-negligible interval of time during which messages (from clients or other replicas) are lost, and thus a recovering replica behaves as being crashed during that time. The unavailability of individual replicas may provoke the unavailability of the overall replicated system. Consider for instance a Byzantine-fault-tolerant replicated system with 4 replicas, and thus able to tolerate one faulty replica (see Section 2.2). If this system is enhanced with recovery mechanisms, then its availability will be affected if (i.) more than one replica recovers at the same time, or (ii.) one replica is recovering while another one is compromised. In both cases, the sum of arbitrary faulty replicas and purposely crashed (recovering) replicas is greater than \( f \), and thus the replicated system will become unavailable during those periods.

This means that recovery mechanisms cannot be added in a straightforward manner to existing Byzantine-fault-tolerant replicated systems. In order to avoid unavailability, recoveries need to be coordinated in an agreed schedule, and the system needs a higher total number of replicas to tolerate both Byzantine (arbitrary) faults and recovery-induced (crash) faults.

2.5 Recoveries Should Introduce Diversity

If a recovery procedure restores a replica to a previous known-to-be-good state (e.g., by rebooting the system from clean media), this replica is still vulnerable to whatever caused its compromise in the first place. Therefore, it is not enough to simply restore replicas to known states, since this would allow an attacker to exploit the same vulnerabilities as before. To address this issue, we argue that the recovery procedure should itself introduce some degree of diversity to restored replicas, so that attackers will have to find other vulnerabilities in order to compromise a replica.

3. THE FOREVER SERVICE

In this section we explain how FOREVER tackles the challenges presented in Section 2.

3.1 Recovery and Evolution

We envisage FOREVER as a service that can be used to enhance the resilience of intrusion-tolerant replicated systems by allowing these systems to tolerate an arbitrary number of faults. Such an ambitious goal is achieved through the
3.3 Coordination

Recovery. FOREVER allows an intrusion-tolerant system to recover from past malicious actions/faults, by cleaning the effects of such actions through (i.) time-triggered periodic recoveries, and (ii.) event-triggered reactive recoveries when malicious behavior is detected. When more than $f$ replicas detect that a given replica $R_i$ is unequivocally faulty (e.g., $R_i$ sends contradictory signed messages in a distributed protocol), this replica is immediately recovered (this is an event-triggered recovery). If, on the other hand, the replicas only suspect that $R_i$ is faulty (but they cannot be certain), recovery of $R_i$ is scheduled in such a way that overall system availability is not affected. Also, time-triggered recoveries are performed periodically in order to neutralize the effects of any undetected faults and intrusions.

Evolution. When FOREVER triggers a recovery of a certain replica, it modifies the vulnerabilities that may be exploited by a malicious adversary (e.g., OS access passwords, open ports, authentication methods). Moreover, FOREVER coordinates online upgrades (e.g., installation of security patches, installation of a new version of the intrusion-tolerant application being protected) with a minimum impact on availability.

These characteristics make FOREVER different from any other recovery service proposed in the past [6, 30, 25].

3.2 Hybrid Model and Architecture

In order to avoid that FOREVER itself is not victim of malicious attacks, an intrusion-tolerant system enhanced with FOREVER should be built under a hybrid system model and architecture [28] in which the system is composed of two parts, with distinct properties and assumptions. These two parts are typically called payload and wormhole. The intrusion-tolerant application (and replication library) runs in the payload part exposed to arbitrary faults and asynchrony. The FOREVER service runs in the wormhole part that is guaranteed to be secure and timely by construction. Figure 1 depicts an intrusion-tolerant application and FOREVER deployed under a hybrid architecture.

In more practical terms, FOREVER should be deployed as a secure real-time subsystem separated from the intrusion-tolerant application being protected. The separation can be achieved through either hardware (e.g., a different PC or a PC appliance board [15, 27]) or software (e.g., using virtualization [2]). Moreover, as it will be explained in Section 3.3, the local FOREVER subsystems running in each replica need to be connected by a synchronous and secure control network.

Although such an hybrid architecture is sufficient to guarantee the security of the FOREVER service, it is not sufficient to guarantee the security and timeliness of the entire recovery process. In fact, FOREVER needs to perform actions over the intrusion-tolerant application, namely it needs to shutdown, recover, and reboot the operating system running in the payload. This means that none of these actions can be allowed to be affected by a malicious attacker. In Section 4 we explain how virtualization can be used to achieve this goal.

3.3 Coordination

Figure 1: FOREVER under a hybrid architecture.

In order to guarantee availability, FOREVER needs to coordinate the recoveries of different replicas. This coordination requires three services:

- **Perfect Failure Detection:** This service is employed to detect if application replicas are alive. In fact each FOREVER monitor monitors each other and, when one is detected to be faulty, the whole replica is considered to be faulty\(^1\).

- **Clock Synchronization:** All local wormholes have synchronized clocks to ensure that the recovery time schedule for each replica is precise.

- **Total Order Multicast:** The algorithm employed to schedule reactive recoveries due to detections [25] uses total order multicast to update the same recovery schedule on every FOREVER monitor without affecting the availability of the intrusion-tolerant application.

Notice that the implementation of these services requires strong timing assumptions, so the FOREVER wormhole subsystem needs to be synchronous.

3.4 Introducing Diversity

One of the most innovative features of FOREVER is the automatic introduction of diversity after a recovery. The idea is that a recovered system should be different from its previous incarnation. The easiest way to do this is to have a pool of operating system images and to start a different one each time the system is recovered. Contrary to common belief, some recent studies reported that different COTS software does not have the same vulnerabilities [10], at least for database management systems. We believe this kind of independence also holds for operating systems.

\(^1\)Since the FOREVER subsystem is faulty on this replica, it cannot be recovered without an administrator manual intervention.
Besides the use of different operating systems and different versions of the same operating system, we intend to use two other features to generate diversity on recovered replicas:

- **Address Obfuscation**: Attacks which exploit memory programming errors (e.g., buffer overflows) are one of today’s most serious security threats. These attacks require an attacker to have an in-depth understanding of the internal details of the system being attacked, including the locations of critical data and/or code. Address obfuscation techniques randomize the location of program data and code each time a program is executed. It has been shown that address obfuscation can greatly reduce the probability of successful attacks [4, 3, 8, 22, 29, 23]. Given that recoveries force the restart of every program in the system, they are a perfect way of introducing address obfuscation both at the OS and application level.

- **System Transformations**: A system transformation \( \tau : I, F, P \rightarrow I', U \) is a function that takes a file system image \( I \) and generates a new file system image \( I' \) updating a set of features \( F \) according to a set of parameters \( P \). The idea of these transformations is to generate new images changing critical parameters such as access ports, disk layouts, usernames, passwords, and permissions; the set of updates performed by the transformation (e.g., the new passwords) is stored in \( U \), to ensure that the new image can be accessed by its authorized users.

We are still beginning to investigate the use of diversity on the FOREVER architecture. A fundamental aspect that needs further investigation is how to generate sufficient diversity while maintaining the compatibility with the intrusion-tolerant application.

### 4. PROTOTYPE

We have already developed a prototype in which the different parts of the system (intrusion-tolerant application and FOREVER) are separated through the use of virtualization.

Virtual machine (VM) kernels, such as the XEN virtual machine monitor [2], may host multiple guest operating systems, every one executed within an isolated VM. In the case of XEN, VMs are called domains. XEN domains do not have all the same privileges, namely there is a special domain with more privileges, called dom0, which is created automatically when the physical machine boots. Domain dom0 builds other domains (called dom1, dom2, etc.). It also performs administrative tasks such as suspending and resuming other VMs, and it can be configured to execute with higher priority than the remaining VMs. These characteristics make XEN an appropriate deployment environment for an intrusion-tolerant application enhanced with FOREVER: the latter would be deployed in domain dom0, while the former would run in a non-privileged domain.

Figure 2 depicts how our current prototype is deployed in each replica. The FOREVER local monitor has access to a file-system repository (FSR) that contains a set of different file-system images (FSI) composed of different operating systems configured with different passwords. Each recovery boots a different FSI and consequently the set of vulnerabilities is changed between recoveries.

Another benefit of using virtual machines is that they allow the recovery process to be more efficient [24]. In our context, FOREVER may instantiate a new payload image (i.e., boot a new virtual replica) before shutting down the old image. If the intrusion-tolerant application is stateless, the transition from the old image to the new recovered image can be almost instantaneous.

### 5. RELATED WORK

FOREVER is an extension of previous work on proactive and reactive recovery [25]. It intends to be a generic recovery and evolution service supporting not only proactive and reactive recoveries but also application and OS online upgrades.

Rejuvenation [13, 9] has been proposed in the 90s as a proactive technique to deal with transient failures due to software aging. The idea is to periodically rejuvenate some software components to eliminate and prevent failures. Some works on this field, beginning with [9], developed techniques to choose the optimal rejuvenation period in order to minimize the downtime of the system. None of the works on software rejuvenation assume faults caused by malicious adversaries, so the problem that we deal in this work is much more complicated than software aging.

Several works advocate the use of proactive recovery to make the system tolerant an unbounded number of malicious faults during its lifetime as long as no more that \( f \) faults occur during a bounded time period [6, 30]. Recently, some of the authors showed that all these works have some hidden problems that could be exploited by a smart adversary [26]. The main problem is that these works assume the asynchronous distributed system model and, under this model it is not possible to guarantee that recoveries are triggered and executed within known time bounds: an adversary can delay the execution of the recoveries and be able to corrupt more...
than $f$ nodes of a system. To make this problem worse, it is impossible to detect such timing attacks under the asynchronous model. The architecture presented in the present paper is based on a hybrid distributed system model [28] and thus uses some trusted and timely components to ensure that replicas are always rejuvenated in accordance to the predefined time bounds.

There is a large body of research that aims to reactively recover systems as fast and efficiently as possible (e.g., [12, 21, 14]). Early work on this field ([e.g., [12]] advocates the execution of some kind of recovery action (in general, restart the monitored process) after a fault is detected. More recently, the recovery oriented computing project proposed a set of methods to make recovery actions more efficient and less costly in terms of time [21]. Several techniques were developed in this project, either to detect failures, restart the minimum set of system components to put the system back to correct operation and to undo some operator configuration errors. These works do not consider Byzantine faults or security-compromised components and also do not rely on redundancy to ensure that the system stays available during recoveries, the main problems addressed by FOREVER.

6. CONCLUSIONS AND FUTURE WORK

Intrusion-tolerant protocols are sufficient to ensure the resilience of replicated systems as long as no more than a certain finite number of replicas is compromised. Unfortunately, this guarantee has limited utility in long-lived systems exposed to malicious attacks. This paper introduced FOREVER, a novel service that can be used to enhance the resilience of intrusion-tolerant replicated systems operating in malicious environments. FOREVER removes faults and intrusions that may happen during system execution, and such removal is done by combining both evolution and recovery techniques.

As future work, we intend to develop novel heuristics to generate diversity after recoveries. Moreover, we intend to assess the benefits and the tradeoff of enhancing a replicated system exposed to malicious attacks with FOREVER. Our ultimate goal is to develop a dependability case of FOREVER and of a system enhanced with this service [1].

7. ACKNOWLEDGMENTS

This work was partially supported by the EC through project IST-2004-27513 (CRUTIAL) and NoE IST-4-026764 NOE (RESIST/FOREVER), and by the FCT, through the Multiannual and the CNU-Portugal Programmes.

8. REFERENCES


