Heterogeneous Networking: A New Survivability Paradigm

Yongguang Zhang
HRL Laboratories, LLC.
3011 Malibu Canyon Road
Malibu, CA 90265, USA
+1-310-317-5147
ygz@hrl.com

Son K. Dao
HRL Laboratories, LLC.
3011 Malibu Canyon Road
Malibu, CA 90265, USA
+1-310-317-5682
skdao@hrl.com

Harrick Vin
The University of Texas at Austin
Department of Computer Sciences
Austin, TX 78712, USA
+1-512- 471-9732
vin@cs.utexas.edu

Lorenzo Alvisi
The University of Texas at Austin
Department of Computer Sciences
Austin, TX 78712, USA
+1-512- 471-9792
lorenzo@cs.utexas.edu

Wenke Lee
Georgia Institute of Technology
College of Computing
Atlanta, GA 30332, USA
+1- 404-894-3152
wenke@cc.gatech.edu

ABSTRACT

We believe that a network, to be survivable, must be heterogeneous. Just like a species that draws on a small gene pool can succumb to a single environmental threat, so a homogeneous network is vulnerable to a malicious attack that exploits a single weakness common to all of its components. In contrast, in a network in which each critical functionality is provided by a diverse set of protocols and implementations, attacks that focus on a weakness of one such protocol or implementation will not be able to bring down the entire network, even though all elements are not be bulletproof and even if some of components are compromised.

Following this *survivability through heterogeneity* philosophy, we propose a new survivability paradigm, called *heterogeneous networking*, for improving a network's defense capabilities. Rather than following the current trend of converging towards single solutions to provide the desired functionality at every element of the network architecture, this methodology calls for systematically increasing the network's heterogeneity without sacrificing its interoperability.

1. INTRODUCTION

The current trend in networking is towards convergence on a single protocol, software, or technology at each layer of the network's architecture. While this trend towards homogeneity results in improved interoperability and reduced costs, it may pose serious vulnerability to the network as a whole. To draw an analogy from the biological sciences, just like a species that draws on a small gene pool can succumb to a single environmental threat, so a homogeneous network is vulnerable to a malicious attack that exploits a single weakness common to all of its components.

For example, it has been pointed out and again that the continued growth of Microsoft products across a large audience has created an environment where one exploit within a Microsoft product may impact a large number of users worldwide. On the other hand, the reason why the Internet survives the recent several rounds of e-mail attacks (e.g., the love bug) is exactly because of the heterogeneity that we are still having in today's Internet – while the love bug exploits the vulnerability in Outlook, it has no effects on Eudora or Unix e-mail clients. Therefore, it may be intuitive that if more diverse technologies are being deployed in a network and if deployed strategically, the network may be more resilient to orchestrated attacks.

Furthermore, building a network with homogeneous elements run the risk of invalidating some of the assumptions at the very core of using fault-tolerant systems to ensure continuous operations of a network even in the presence of attacks. For instance, techniques developed to tolerate arbitrary (Byzantine) failures have been proposed as a way to make a system survivable to security attacks. The basic idea behind these techniques is to replicate critical components so that, if the number of arbitrarily faulty replicas does not exceed a given threshold *t*, the system will continue to operate correctly. Clearly, critical to the correctness of all these approaches is the determination of an appropriate value for *t*. The chosen value should be such that the probability that at any point in time the number of concurrent failures exceeds *t* is negligible. In classical fault-tolerance literature, this probability is computed assuming that failures are independent: in other words, the failure of a replica does not affect the probability that another replica will also fail. If such fault-tolerance techniques are used to tolerate security attacks in

Conference '00, Month 1-2, 2000, City, State.

a network with homogeneous elements, the assumption of failure independence is ill founded. In other words, for security attacks it is not reasonable to assume that identical replicas will fail independently: rather, once a successful attack is performed against one replica, the same attack can be performed successfully on all identical replicas. To restore the assumption of failure independence, we need to introduce sufficient heterogeneity back to the network.

In this paper, we propose a new paradigm that achieves network survivability through the use of heterogeneous technologies. We propose a network architecture in which each critical functional capability is provided by a diverse set of instantiations or implementations, so that attacks that focus on a weakness of any one such protocol or implementation is less likely to prevent the network from providing acceptable service.

2. HETEROGENEOUS NETWORKING MODEL

Our vision of "*survivability through heterogeneity*" is founded on the observation that different instances of network elements that export the same functional capability are, in general, vulnerable to different security attacks. Hence, a network architecture that supports a collection of heterogeneous network elements for the same functional capability offers a greater possibility of surviving security attacks as compared to homogeneous networks. Consider, for instance, the following two examples.

1. A router is an important element of network architecture. A network with homogeneous routers (and hence homogeneous router operating systems) is more susceptible to security attacks than a network architecture that employs a heterogeneous collection of routers with multiple, redundant paths through heterogeneous routers between every source-destination pair.
2. End-to-end network services rely on transport protocols for reliable, timely delivery of data packets; the survivability of such network services depends critically on the ability of transport protocols to survive attacks. Hence, a web service that can utilize UDP or SRDP (Simple Reliable Datagram Protocol) in addition to TCP for data transport can survive a TCP SYN-flood attack (which is the cause of several denial-of-service attacks on web servers today).

As these examples illustrate, the survivability of a heterogeneous networking framework depends critically on the differences in the vulnerability to security attacks of different instantiation of network elements at each level of functional capability. The greater the diversity in the vulnerability of network elements to attacks, the higher the survivability of the heterogeneous networking framework.

2.1 Diversity Space

Conceptually, we can represent the functional capabilities of network architecture and the heterogeneity of network elements using *diversity space diagram*. This diagram organizes functional capabilities of a network (e.g., network and transport protocols, routing protocols, router operating systems, etc.) into a multi-dimensional space. Each network element that instantiates a functional capability is represented as a point along the dimension. Figure 1 illustrates an example of such diversity space. Here, UDP, RTP and TCP are the three network elements along the dimension of transport protocols, while satellite, wireless, and fiber-optic networks are examples of elements for the communication medium (or physical network connectivity).

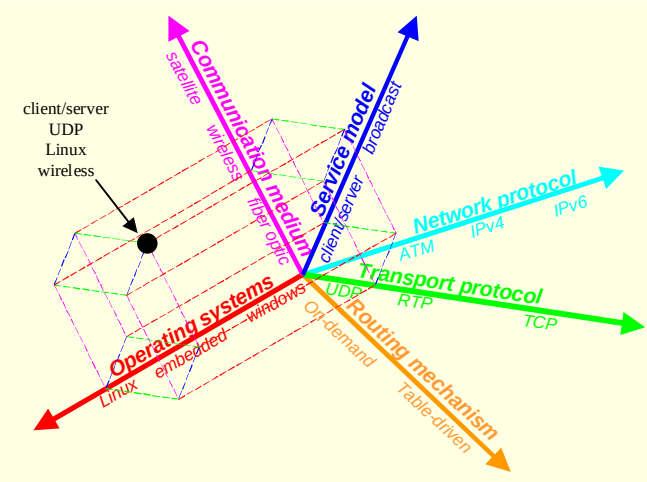

Figure 1 The diversity space for heterogeneous networking

The *distance* between two network elements along any dimension reflects the diversity in their vulnerability to attacks; the larger the distance between two network elements along a dimension, the smaller is the overlap in their vulnerability to attacks. For example, the distance between "Linux" and "Windows" in the operating system dimension is relatively large because these two systems are independently designed and implemented, while the distance between "IPv4" and "IPv6" is relatively small because the latter is derived from the former.

2.2 Vulnerability Model and Survivability Measure

Given such a diversity space diagram, the key question one has to address in designing a survivable network is: for each of the dimensions, which and how many network elements should a survivable network framework support?

This question can be addressed by developing a *vulnerability model* for each network element, and by introducing the novel concept of "survivability measure" – a metric for capturing the diversity in the vulnerability to attacks of different network elements. In particular, we can identify, for each network element, the set of attacks that the network element is vulnerable to. Let A denote the cumulative set of such attacks. Then, a survivable network framework should include, at a minimum, the set S of network elements at each level of functional capability such that at least one network element in S is not vulnerable to each of the attacks in A. In practice, the set S may include network elements such that several network elements are vulnerable to each of the attacks in A. We can then develop a quantifiable *survivability measure* for set S; this measure will capture the extent of redundancy required in S so as to reduce the likelihood that every element in S is vulnerable to an unknown future attack. Intuitively, the higher the survivability measure is,

the more "diverse" the set is. The more "diverse" a network becomes, the more time/resources an adversary must invest to identify vulnerabilities of all elements and to plan orchestrated attacks on each of them.

Our methodology for constructing the survivable set *S* is guided by the following conjecture: *survivability of the network elements in set S to the set of known attacks A is a reasonable indicator of the degree to which set S will survive unknown attacks*.

There may be many ways to define a quantifiable survivability measure for a given set of network elements that export the same functional capability. One measure is the cumulative diversity distance between all pairs of elements in the set. Another measure can be the number of distinct attacks that the set can tolerate.

Once we identify the set of network elements *S* for each level of functional capability, we can design and implement the relevant network elements to create our heterogeneous networking framework. The key challenge is to create a systematic plan for instantiating network elements with reasonable cost and with manageable complexity. A successful instantiation of these network elements will yield a network that will be highly resilient to a vast variety of known an unknown security attacks.

3. DESIGNING HIGHLY SURVIVABLE OVERLAY NETWORKS AND SERVICES

End-to-end services involve layered implementation of functional capabilities; this can be realized through composition of network elements. In our heterogeneous networking framework, each functional capability is instantiated using a set *S* of heterogeneous network elements. Hence, in principle, composing together different selections of network elements from each functional capability layer can yield different versions of an end-to-end network service. For example, Figure 2 depicts a composition of several network elements to create WWW and broadcast services. It is easy to see that the broadcast service can also be instantiated by using RLM (reliable layered multicast) instead of UDP as its transport protocol. In such a framework, the network can support half of the services using RLM and the other half using UDP, or it can utilize one of the two instantiations during normal operation and switch to the other instantiation on detecting an attack.

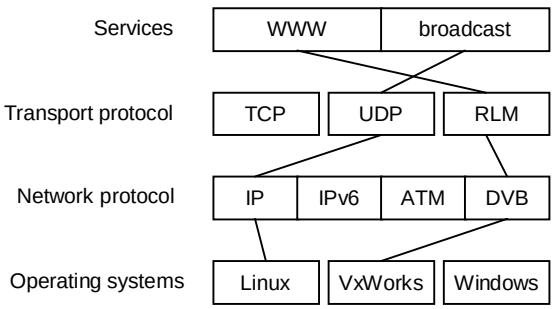

Figure 2 Interchangeable elements at each layer

Realizing this in practice imposes several challenges. This is because not all network elements within a network layer may be functionally equivalent from the perspective of an application. For instance, both TCP and UDP are transport protocols; however, TCP provides to an application a reliable transport with mechanisms for congestion control, while UDP does neither. Hence, even though TCP and UDP belong to the same network layer, it is, in general, impossible to switch among them in a way that is transparent to the application.

This issue can be addressed by the following four mechanisms:

- *Patching lost functionality*. This approach would be to implement any functionality that may be lost while switching from one network element to another at a higher level in the network protocol stack. For instance, on switching from TCP to UDP at the transport protocol level, the functionality of reliable transmission and congestion control can be implemented at the session or higher layer. This approach has the advantage of supporting application transparency, but has the major disadvantage that the resulting implementation may be vulnerable to the same attack of the network element that it is trying to substitute for.
- *Tolerable operation region*. The approach is to renounce transparency, at least partially, and require the application to specify an acceptable region of operation in the heterogeneous diversity space. If an attack merits network service reconfiguration that is outside the application-specified tolerance, then the application is notified through an upcall interface. The application can provide specific handlers to adapt appropriately in response to these upcalls.

  If, on the other hand, the network operates within application-specified tolerance, then any reconfiguration of network service through recomposition of network elements is transparent to the application. To enable such transparent reconfiguration, each network element must export a well-defined interface. Further, the heterogeneous networking framework should export a set of mechanisms to translate and transfer state among network elements providing the same functional capability.
- *Overlay networks*. Using the above two approaches, the heterogeneous networking framework can now support logical overlay networks with multiple physical realizations. Operating such overlay networks also present several design choices. In the simplest case, the framework can use one of the physical realizations as a default, and switch to other realizations only on detecting an attack. In somewhat more complex settings, the framework may simultaneously support multiple physical realizations of the logical overlay network; each physical realization carrying a fraction of the total overlay network traffic. Traffic can be distributed at various levels of granularity: from the packet level to flows to aggregates of flows. These design choices will have implications on the network's ability to support quality of service (QoS) guarantees. This is because, to provide end-to-end service guarantees, a network may need to reserve resources along a path, as well as initialize and maintain state information at each network element. Consequently, switching among different physical realizations on a per-packet basis may violate application's QoS requirements.

- *Multiplexing*. It is quite often that one element in one layer needs to interact with heterogeneous elements of another layer. For example, a WWW server may need to serve clients using TCP or using RLM at the same time. This requires multiplexing techniques to divide one service into multiple forms to be served by heterogeneous alternatives. As another example, a mission critical network can be overlaid on several heterogeneous networks that provide similar connectivity. The overlay mechanism will ensure that it can dynamically change its affiliation with underlying alternatives when one is under attacks.

4. NETWORK RECONSTITUTION THROUGH HETEROGENEOUS REPLICATION

Replication has been used in distributed systems as a fault-tolerance measure. When a system component fails, a replicated component takes over the functionality of the failed component so that the system as a whole can accomplish its mission. As we have pointed out earlier, traditional replication measures in a computer network, such as backup routes or backup servers, can improve the network's resilient against unintentional failures, but will not improve its survivability against orchestrated attacks.

Our heterogeneous networking methodology supports a new type of replication – replication of critical network elements – such as connectivity infrastructure, resources, and services – over heterogeneous components. When a successful attack diminishes the functionality of a network element, a heterogeneous replica of the element may still function as usual. Hence, a network can switch to a different, functionally equivalent network element and continue to provide the same end-to-end service to applications. We refer to this approach to survivability as *network reconstitution through heterogeneous replication*.

This network reconstitution approach consists of the following two basic steps:

- *Heterogeneous replication*. This is to replicate the critical network functional capabilities, not by duplicating the components that export these capabilities, but by instantiating them into many different network elements. This can be done by physically duplicating the network components, and having different network elements activated at each components, or by having more than one network elements co-exist at the same physical component. To develop the mechanisms for heterogeneous replication, we will build upon the tools for (off-line) switching and migrating network elements as described in previous sections
- *Dynamic reconfiguration*. This is to reconfigure, on the fly, the composition of network elements. When an attack seriously damages a functional capability provided by a network element, the system can dynamically switch to a replicate of the same functional capability.

Furthermore, dynamic reconfiguration can be used as a preemptive measure. By frequently changing the active set of elements, the network may have taken away the ability for an adversary to identify weaknesses and time needed to plan for an orchestrated attack.

Our network reconstitution techniques are also built upon the following:

- A set of policies that define what critical elements in a network we should replicate, what type of heterogeneous components we should replicate onto, and how to coordinate between replicas during normal operations and during attacks.
- Mechanisms that mediate between intrusion detection algorithms and our heterogeneous networking platform, so that any attack detected by the intrusion detection module can trigger dynamic reconfiguration actions.

One important issue we need to address is to identify as to what we should replicate and what type of heterogeneous replications do we need. We will address this issue through the threat model and the additional intrusion detection component in the next section.

5. THE ROLE OF INTRUSION DETECTION

We will introduce an intrusion detection component in our new survivable network paradigm as an optimization measure. The role of intrusion detection here is to recognize the threats to network services and to provide information about the attacks so that appropriate recovery actions can be carried out. Threat models of network, which specify the essential services and their degrees of tolerable performance degradation or damage, are used by the intrusion detection system (IDS) to determine what to monitor and what constitutes threats. Reports of detected threats by the IDS describing the compromised services and attack techniques are then used to determine which heterogeneous replications should be activated.

In a survivable network where the mission must be fulfilled in a timely manner in the presence of attacks, a threat is an attack scenario that aims to compromise or damage the *essential components/services*. Attacks targeted at nonessential services need not be considered as threats and thus do not warrant network recovery actions, especially when there is limited response time and resources, which is normally the case when the network is under orchestrated attacks.

A *threat model* formally specifies, for a specific mission (i.e., normal usage scenario), which network component/service is critical and which isn't, and for each of these components/services their acceptable quality requirement (or its degree of tolerable performance degradation or damage). The threat models link the policies/requirements with survivability mechanisms because they enable the *recognition* of on-going threats to the network and its mission, and hence facilitate the decision-making on when and how the heterogeneous replicas can be used to recover and reconstitute the mission. As an example, consider a WWW server. Its threat model includes: essential service – to provide information of upon request, minimum quality requirement – to service at least x number of concurrent requests with at most y seconds of delay. This model dictates that if the service is not up to the performance requirement, it is a threat and recovery action must be taken to recover the service.

Because there can be potentially a large number of threats, we can introduce the notion of threat taxonomy where similar threats can be grouped together. The taxonomy can reduce the system complexities because it not only provides a common terminology for referring to the threats but also allows the same recognition and recovery techniques be applied to the same category of threats. For example, we can use the following three dimensions to categorize threats: the *effect* (or goal), e.g., denial-of-service; the *target*, i.e., which essential service is targeted; and the *technique*, i.e., how is the threat carried out. For example, denial-of-service (DoS) can be accomplished by two techniques: "crashing" the server or "resource consumption". Two threats are in the same category if they have the same values in all three dimensions.

In our architecture, the intrusion detection component can list the detected on-going threats and the predicted upcoming threats, based on attack scenario analysis. Using information of the threats, i.e., the effects, targets, and techniques, appropriate recovery actions can be carried out. In particular, the technique dimension determines what type of heterogeneous replication should be used, i.e., *how* to use the heterogeneous replications, for the damaged service(s). For example, if a DoS attack is accomplished via exploiting a bug in Windows and causing the server to crash, then a Linux implementation can be activated. If the DoS attack is accomplished via exploiting TCP handshake (e.g., it is a SYN-flood attack), then other implementations using other transport layer protocol can be activated. To generalize the solution, the threat techniques should be mapped to dimensions of Diversity Space (see Section 2.1) and a heterogeneous replication should be selected automatically so that it has the longest distance from the one that was subject to the identified threat.

6. HETEROGENEOUS SERVICE MODEL

In this section, we will further demonstrate the power of our new survivable network paradigm, using an example heterogeneous service model. This example is implementable and illustrates the benefits of the new ideas explained in this paper.

The current Internet service model is rather homogeneous; many applications have been converging towards the WWW browser/server model. While the standardization on WWW model may have saved costs, the WWW client/server model does have a fundamental weakness – it is often subjected to distributed denial of service (DDoS) attacks, where an adversary through controlling large number of unsuspected clients launches illegitimate or seemingly legitimate but useless requests so overwhelming as to deny truly legitimate clients a chance to be served. This is especially so in a network with symmetric bandwidth, such as the Internet core. Since the WWW service model is often asymmetric in bandwidth requirements (more data flowing from servers to clients than in the other direction), broadband connectivity may have an unintended negative effect: the idle bandwidth in the direction from clients to servers could make DDoS attacks more effective. For all these reasons, today's WWW services are still largely defenseless in front of DDoS attacks.

Having an asymmetric network infrastructure may help restrain the DDoS attacker. If the available bandwidth in one direction (from clients to servers) can be limited without affecting the other direction (from servers to clients), we can take away the resources that fuel the DDoS attacks. And the satellite networks may be a prefect fit for this purpose, because the bandwidth disparity between the two directions, downlink from server to client and uplink from client to server, is often as large as 10000 times. For example, the downlink in next generation satellite networks will be typically 100Mbps, but the uptime will usually limited to 128Kbps or 512Kbps. Therefore, DDoS will not be as effective in a highly asymmetric network like satellite networks.

Furthermore, such attacks may be completely useless, if a different service model is used. For example, broadcast-based information dissemination service can be used to provide WWW server, in which the servers actively broadcast the information to all clients and the clients passively receive all data and then selectively filter out the useful information. This model suits especially well for applications where information flows are highly asymmetric and works effectively over satellite networks [1]. More importantly, broadcast-based information dissemination model is immune to DDoS attacks because it does not operate on user requests.

Therefore, we can apply our proposed "heterogeneous networking" paradigm and build a survivable network application on two completely different sets of service models and over two different network infrastructures (see Figure 3). When one service is degraded significantly due to attacks on one or more elements involved, the application can quickly migrate to the second service.

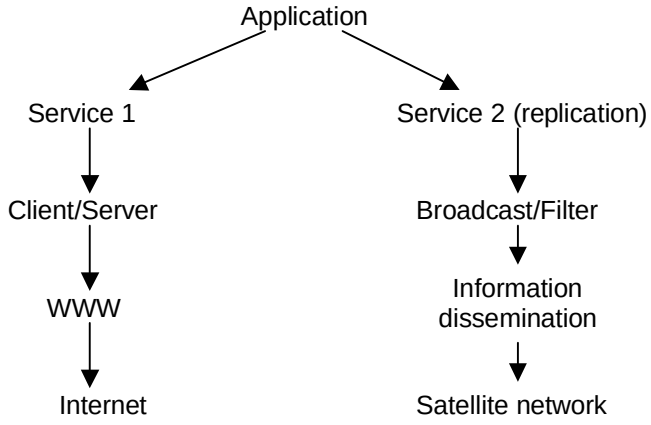

Figure 3 Replications over heterogeneous service models

7. CONCLUSION

The philosophy of survivability through heterogeneity has long been a fascinating idea. For example, in a report published in 1999, CERT had suggested that one possible technique for recovery of essential services after attack is to use redundant modules with the same interface but different implementation [2]. Several DARPA Information Survivability projects, e.g., the Immunix project by OGI [3], also listed "heterogeneity" (different implementation from the same specification) as one of the main objectives. Heterogeneity has also been exploited to try to achieve tolerance from software faults through n-versions

programming [4]. However, to the best of our knowledge there still hasn't been any success in terms of actual design and implementation example of the "survivability through heterogeneity" principle. We believe our ideas of putting this philosophy at work through our heterogeneous networking paradigm are truly unprecedented.

8. REFERENCES

[1] E. C. Shek, S. K. Dao, Y. Zhang, D. J. Van Buer, and G. Giuffrida, "Intelligent Information Dissemination Services in Hybrid Satellite-Wireless Networks," in ACM Mobile Networks and Applications (MONET) Journal, Vol 5, No 4, pp. 273-284, December 2000.

[2] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson, T. A. Longstaff, and N. R. Mead, "Survivability: Protecting Your Critical Systems," in IEEE Internet Computing, November/December 1999.

[3] C. Cowan and C. Pu, "Immunix: Survivability Through Specialization," in Proceedings of SEI Information Survivability Workshop, San Diego, California, February 1997.

[4] A. Avizienis. "The N-Version Aproach to Fault-Tolerant Software," in IEEE Transactions on Software Engineering, Vol. SE-13, pp. 1491-1501, May 1987.