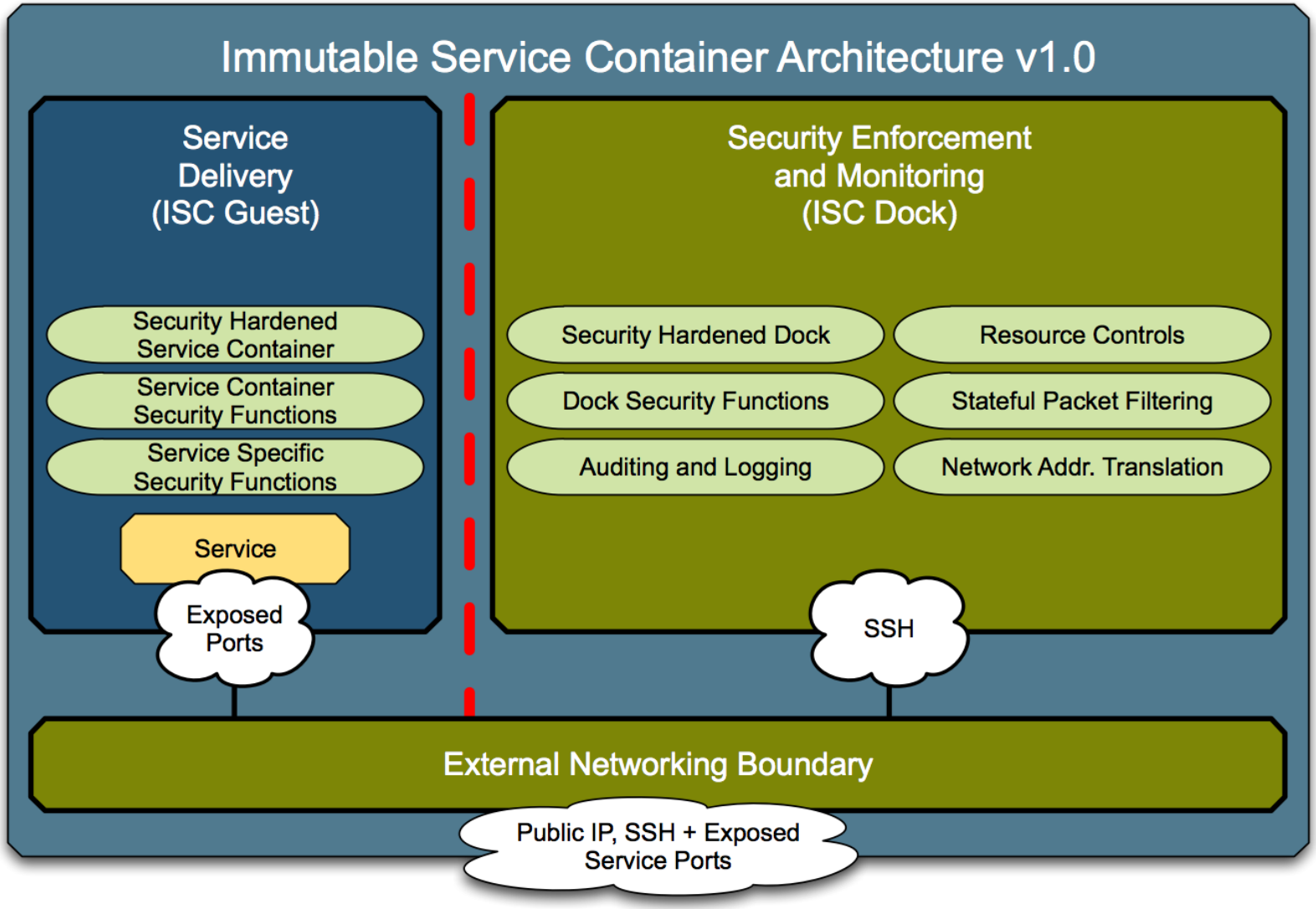
Immutable Service Containers

Glenn Brunette
Distinguished Engineer
Chief Security Architect
Sun Microsystems, Inc.

What are they?

An Immutable Service Container (ISC) is architectural pattern with associated deployment strategies defining a new foundation for highly secure service delivery.

What do they look like?



Why are they interesting?

- Best-in-class security configurations
  - Built upon industry accepted recommended security practices.
  - Not an expert? Leverage pre-integrated reference configurations!
- Common configurations, unlimited uses
  - Architecture is not specific to Clouds, DMZs, or other use cases.
  - Supports virtually unlimited applications and service delivery models.
- Consistent, repeatable packaging, deployment, management
  - Foundation for secure, service-driven “golden images”.
  - Build once, deploy everywhere, replace on update.
  - One “service” installed and exposed per ISC.

Why should you care?

- For developers and application owners:
  - ISCs help to protect applications and services from tampering
  - ISCs provide a consistent set of security interfaces and resources for applications and services to use
- For system administrators:
  - ISCs isolate services from one another to avoid contamination
  - ISCs separate service delivery from security enforcement/monitoring
  - ISCs can be (mostly) pre-configured by security experts
- For IT managers:
  - ISC creation can be automated, pre-integrating security functionality and making them faster and easier to build and deploy
  - ISCs leverage industry accepted security practices making them easier to audit and support

OpenSolaris ISC Capabilities

Pre-Integrated Security Protection #1

Minimized Network Attack Surface

- **Packet Filtering**
  - Enabled automatically with a default deny configuration.
  - SSH & DHCP are the only default exceptions (for ease of use).
  - Policy can be adjusted based upon local requirements.
  - Access logs are published to syslog (default: `/var/log/ipflog`).
- **Non-Executable Stack**
  - Most binaries linked against `noexstk.map` file.
  - Kernel enforcement enabled via `noexec_user_stack` setting.
  - Assists in the mitigation of certain buffer overflow attacks.
- **Secure by Default + OS Hardening + (Opt.) Minimization**
  - Implemented Center for Internet Security guidelines (adapted).

Pre-Integrated Security Protection #2

Enhanced Protection of Sensitive Content

- **Encrypted Network Communications**
  - Secure Shell is the default and only method of access.
  - IPsec/IKE can be optionally configured/enabled as needed.
- **Encrypted Swap Space**
  - Ephemeral AES-256 keys generated from `/dev/random`.
  - Useful for protecting sensitive content that is “swapped out”.
- **Encrypted Temporary (Scratch) Space**
  - Ephemeral AES-256 keys generated from `/dev/random`.
  - Configurable location and size for encrypted scratch space.
  - Useful for storing sensitive information (e.g., key material).

Pre-Integrated Security Protection #3

Improved Isolation of Services from Management

- **Non-Global Zone**
  - Mandatory isolation between services running in different zones
  - Prohibited from direct access to kernel memory, devices, etc.
  - Entire environment operates with reduced privileges
  - Restricted access provides protection against certain root kit methods
  - Flexible file system policy enabling selective file system immutability
  - Dynamic resource control policy (for CPU, memory, etc.)

Detailed discussion of the security capabilities of non-global zones is found in the Sun BluePrints article titled "Understanding the Security Capabilities of Solaris Zones Software", http://mapping.sun.com/profile/offer.jsp?id=120

Pre-Integrated Security Protection #4

Improved Isolation of Services from Management

- **Private Virtual Network**
  - Every non-global zone is assigned its own virtual NIC and IP address
  - Non-global zone external connectivity fails if IP address is changed
  - Private Virtual Network is not directly reachable from external hosts
- **Network Address Translation / Port Address Translation**
  - Gatekeeper for inbound and outbound access to non-global zones
  - All outbound access is permitted by default (for ease of use)
  - No inbound access is permitted by default (for security)
  - Policy can be easily adjusted based on specific requirements

Pre-Integrated Security Protection #5

Observability with Integrity

- **Operating Environment Auditing**
  - Enabled by default for the global and all non-global zones
  - Auditing configuration and logs controlled from the global zone
  - Individual non-global zones cannot access or modify audit trail
  - Login and logout events, administrative events, executed commands (with command line arguments)
  - Records published to binary audit trail and syslog (`/var/log/auditlog`)

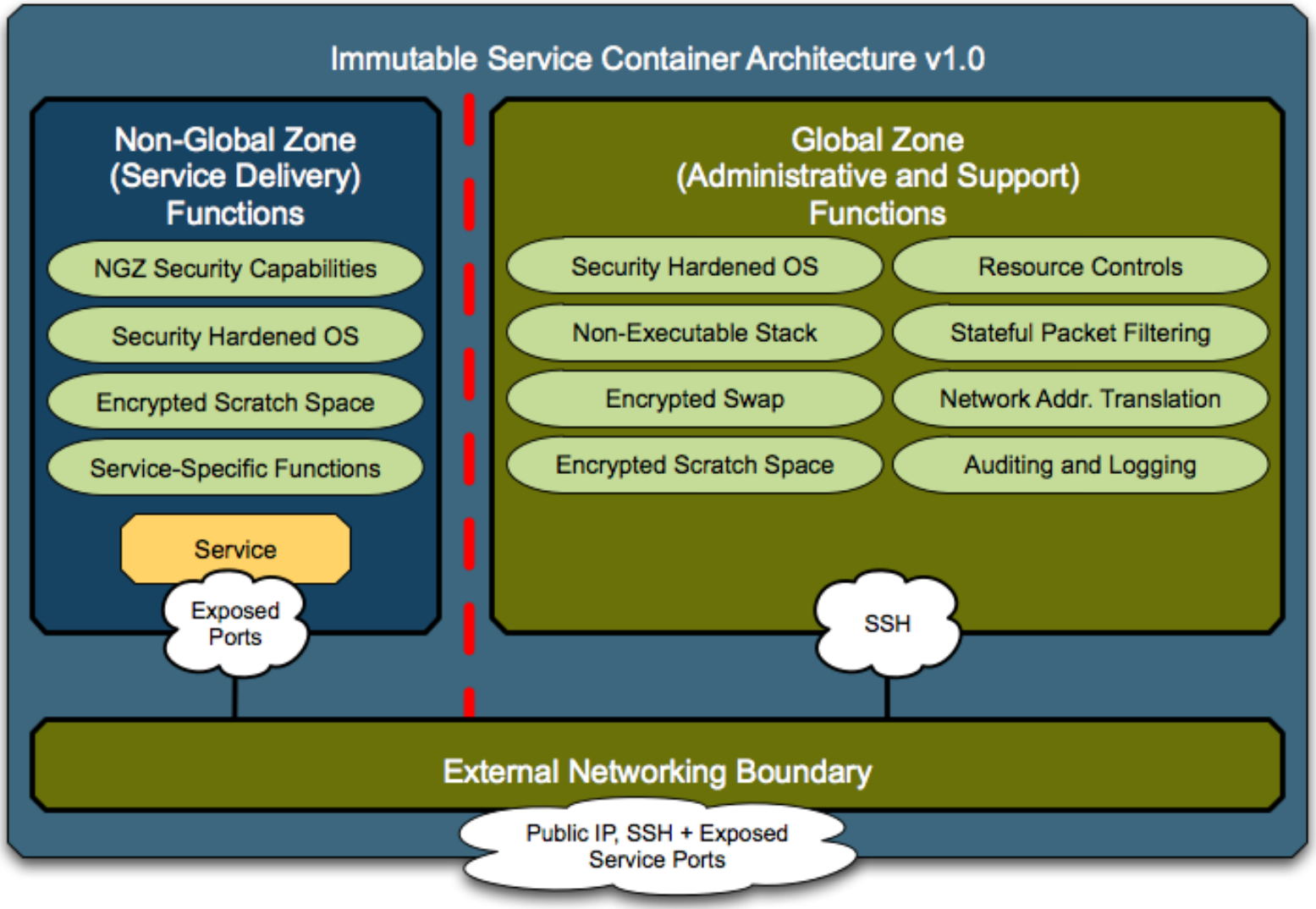
Building an OpenSolaris ISC

Implementing a Basic ISC Architecture

- Install OpenSolaris 2009.06
  - Get it now! http://www.opensolaris.com/
- Download and Install the ISC Construction Kit Preview
  ```
  $ hg clone https://kenai.com/hg/isc~source  isc
  ```
- Execute the ISC Constructor (to create the Dock & ISC #1)
  ```
  $ pfexec isc/bin/iscadm.ksh -c -d -i -n 1
  ```
- Update boot archive and restart the system to complete the changes.
  ```
  $ pfexec bootadm update-archive
  $ pfexec shutdown -g 0 -i 0 -y
  ```

Detailed instructions can be found on the Immutable Service Container Kenai project page, http://kenai.com/projects/isc/pages/Home

OpenSolaris ISC Architectural Diagram



Adding New Services Example (Apache)

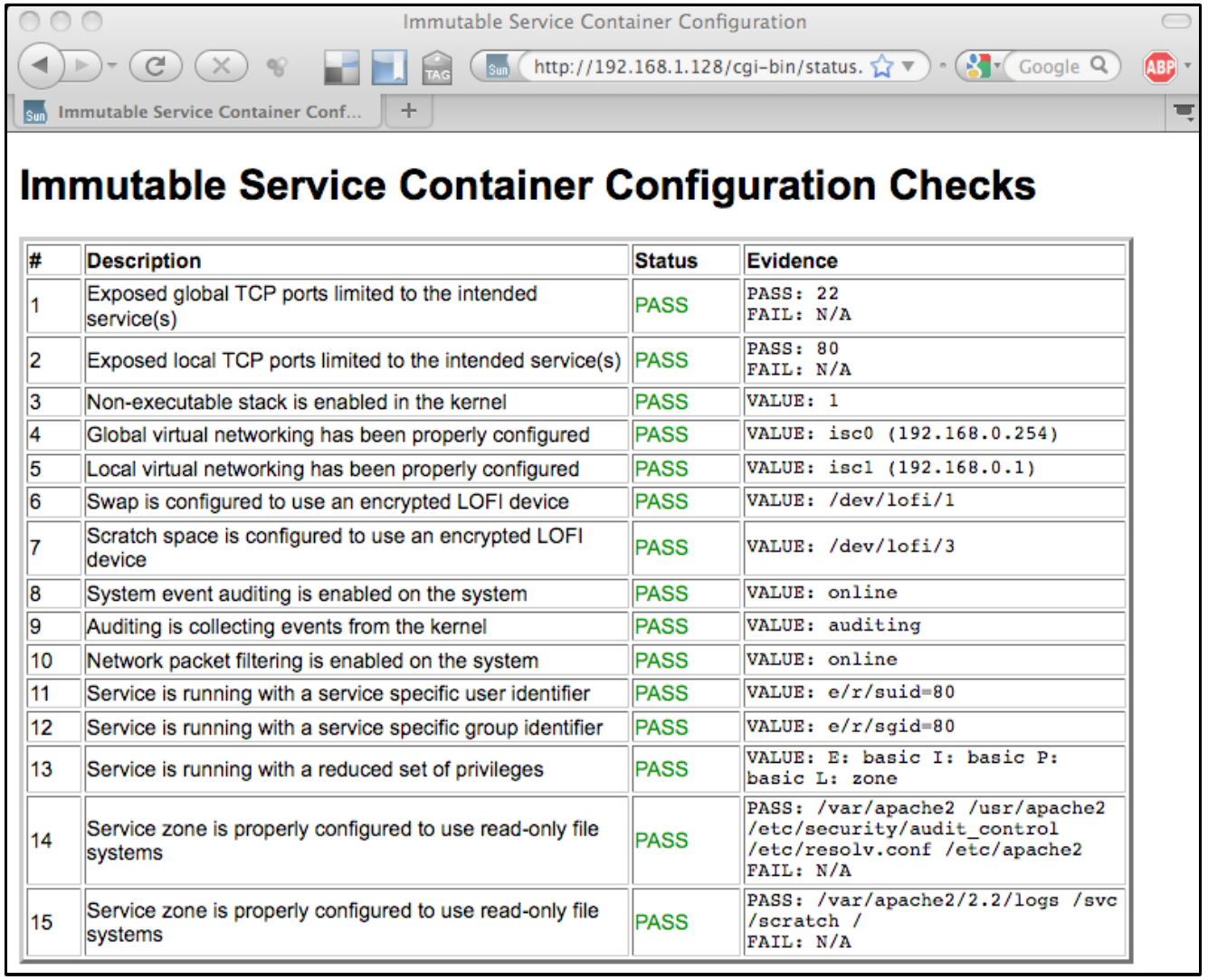
- Non-Global Zone Operations:
  - Install the new service (if necessary):
    ```
    $ pfexec pkg install SUNWapch22
    ```
  - Configure and enable the new service:
    ```
    $ pfexec svcadm enable apache22
    ```
- Global Zone Operations:
  - Adjust the IP NAT policy (/etc/ipf/ipnat.conf):
    ```
    rdr e1000g0 0.0.0.0/0 port 80 -> 192.168.0.1 port 80
    ```
  - Adjust the IP Filter policy (/etc/ipf/ipf.conf):
    ```
    pass in quick on e1000g0 proto tcp from any to
    192.168.0.1 port = 80 keep state
    ```
  - Apply the new IP Filter and IP NAT policies
    ```
    $ pfexec ipnat -FC -f /etc/ipf/ipnat.conf
    $ pfexec ipf -Fa -f /etc/ipf/ipf.conf
    ```

Post-Installation of New Services

- Non-Global Zone Operations:
  - > Use unique credentials and least (process) privileges
  - > Optimize the security configuration of the new service
  - > Leverage encrypted scratch space for sensitive content
  - > Consider using a custom file system layout for immutability
    - – Read only: binaries, libraries, etc.
    - – Read write: logs, configuration files and data (if needed)
- Global Zone Operations:
  - > Implement resource controls for the service non-global zone

Note: These steps can all be fully automated along with the service installation to promote consistent deployments for all "like" services.

OpenSolaris ISC Validation



Immutable Service Container Configuration Checks

#	Description	Status	Evidence
1	Exposed global TCP ports limited to the intended service(s)	PASS	PASS: 22 FAIL: N/A
2	Exposed local TCP ports limited to the intended service(s)	PASS	PASS: 80 FAIL: N/A
3	Non-executable stack is enabled in the kernel	PASS	VALUE: 1
4	Global virtual networking has been properly configured	PASS	VALUE: isc0 (192.168.0.254)
5	Local virtual networking has been properly configured	PASS	VALUE: isc1 (192.168.0.1)
6	Swap is configured to use an encrypted LOFI device	PASS	VALUE: /dev/lofi/1
7	Scratch space is configured to use an encrypted LOFI device	PASS	VALUE: /dev/lofi/3
8	System event auditing is enabled on the system	PASS	VALUE: online
9	Auditing is collecting events from the kernel	PASS	VALUE: auditing
10	Network packet filtering is enabled on the system	PASS	VALUE: online
11	Service is running with a service specific user identifier	PASS	VALUE: e/r/suid=80
12	Service is running with a service specific group identifier	PASS	VALUE: e/r/sgid=80
13	Service is running with a reduced set of privileges	PASS	VALUE: E: basic I: basic P: basic L: zone
14	Service zone is properly configured to use read-only file systems	PASS	PASS: /var/apache2 /usr/apache2 /etc/security/audit_control /etc/resolv.conf /etc/apache2 FAIL: N/A
15	Service zone is properly configured to use read-only file systems	PASS	PASS: /var/apache2/2.2/logs /svc /scratch / FAIL: N/A

Don't want to build your own?

- Pre-installed Open Virtualization Format (OVF) Images
  - http://kenai.com/projects/isc/pages/OpenSolaris
- Pre-installed Amazon EC2 Machine Images (AMI)
  - http://blogs.sun.com/ec2/
  - ami-48c32021 (US), ami-78567d0c (EU)
- also portions of the OpenSolaris ISC codebase are used by:
  - Security-Enhanced Amazon EC2 Machine Images:
    - OpenSolaris 2008.11 (US, EU)
    - OpenSolaris 2009.06 (US, EU)
    - OpenSolaris 2008.11 + AMP + Drupal (US, EU)
  - OpenSolaris 2009.06 JeOS Prototype Images:
    - http://hub.opensolaris.org/bin/view/Project+jeos/WebHome

Future Direction

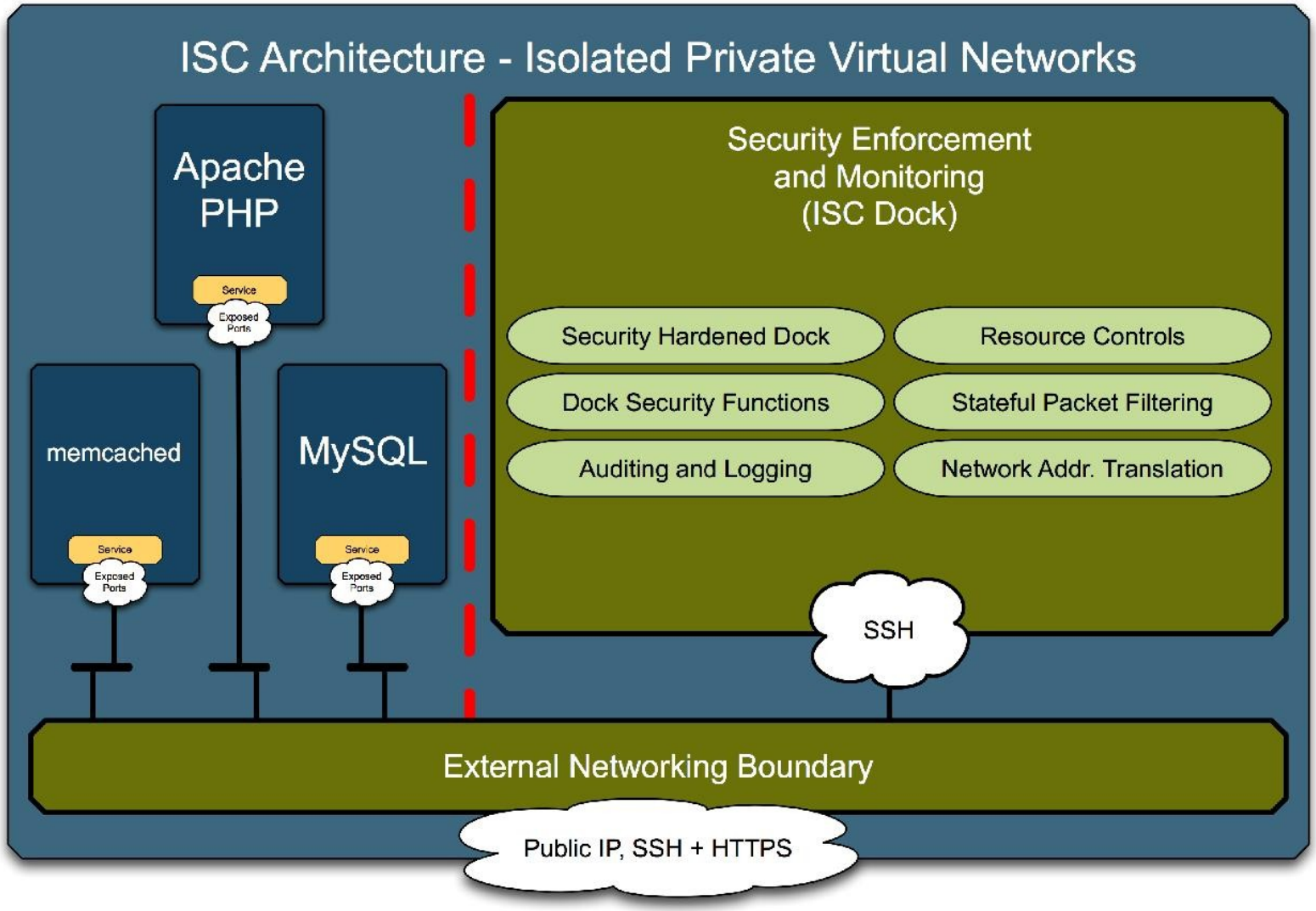
OpenSolaris ISC Next Steps

- Integration with emerging OpenSolaris capabilities:
  - ZFS Crypto (for Swap Space, Scratch Space, and Storage)
  - Anti-spoofing Link Protection (Crossbow v1.3)
  - Always On Auditing
  - Just Enough OS (JeOS)

Other Ideas to Consider

- Automation of IP Packet Filtering and NAT Operations
- Automation of Periodic Forensic Snapshots
- Automation of Basic File and Network Monitoring
- Support for Additional ISC Models (beyond Non-Global Zones)
- What else? Solaris 10 version? Virtual Box version?

New Private Virtual Network Models



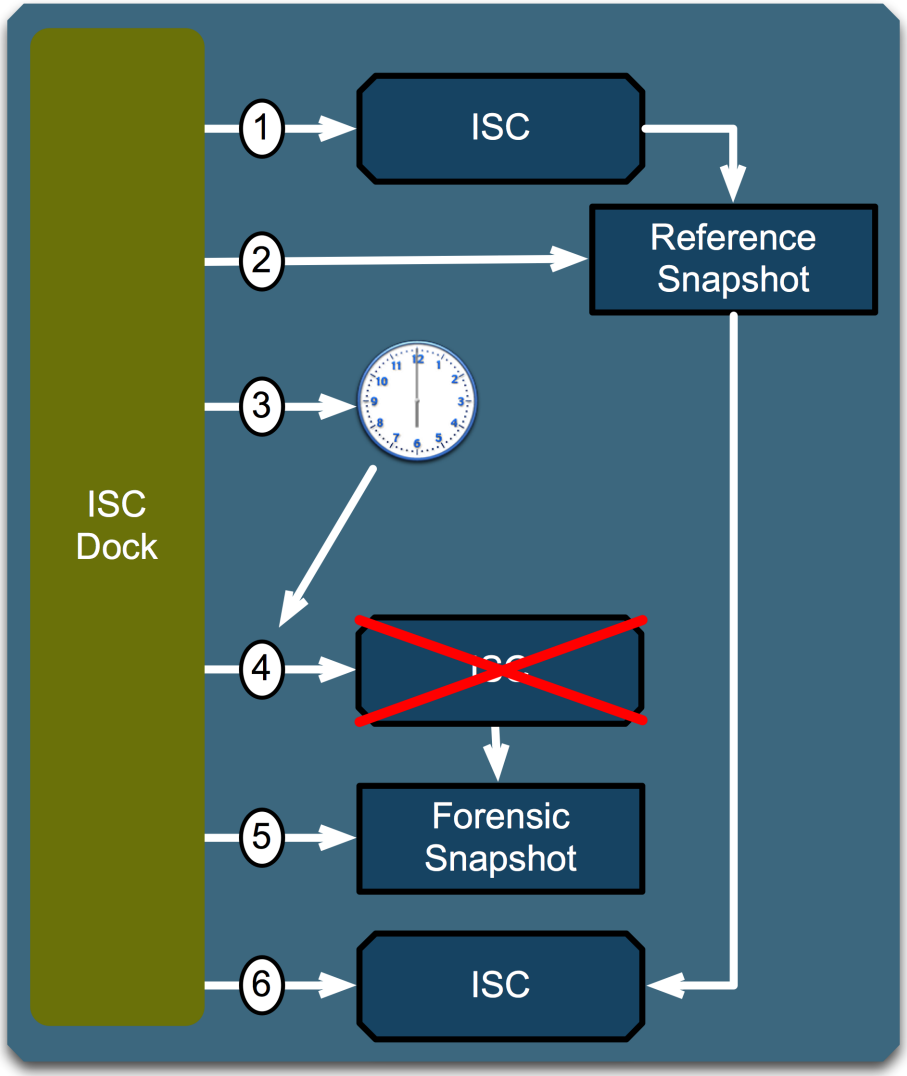
Autonomic ISC Self-Cleansing

Preparation:

1. ISC is created using the standard process running on the ISC Dock.
2. The ISC Dock creates a Reference Snapshot of the ISC. This snapshot serves as the baseline for self-cleansing (rollback) operations.
3. A timer is enabled and set to expire at maximum lifetime of the ISC.

Operation:

4. Upon the expiration of the maximum lifetime timer, the ISC Dock instructs the ISC to gracefully shutdown. The ISC may also perform other support tasks associated with this operation.
5. Once halted, the ISC Dock creates a Forensic Snapshot of the ISC for later analysis (if desired).
6. ISC is rolled back to the Reference Snapshot, the ISC is re-started, and the maximum lifetime timer is reset.



References

- Immutable Service Containers (General, Architecture, Networking, etc.):
  - http://kenai.com/projects/isc/pages/Home
- Immutable Service Container Podcasts:
  - Innovation(at)Sun:
    http://blogs.sun.com/innovation/entry/immutable_service_containers
  - HELDENFunk:
    http://mediacast.sun.com/users/constant/media/HELDENFunk_030_20090403.mp3
- OpenSolaris-based Immutable Service Containers:
  - http://kenai.com/projects/isc/pages/OpenSolaris

Immutable Service Containers

Glenn Brunette
glenn.brunette@sun.com
http://blogs.sun.com/gbrunett
LinkedIn/Twitter/Facebook/Flickr Key: gbrunett