Implementing SHIM6 using the Linux XFRM framework

Sébastien Barré, Olivier Bonaventure

Université catholique de Louvain
`http://inl.info.ucl.ac.be`

Dec. 14th, 2007

Routing in Next Generation Workshop

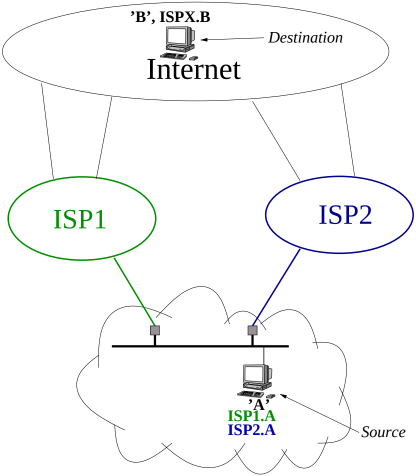
1. Introduction
   - Introduction to Shim6
   - Shim6 : a new layer
   - The REAP exploration protocol
2. The life of an Internet communication...
   - When the TCP SYN is sent...
   - XFRM comes into play
   - Detecting failures : REAP
   - Garbage collection
3. Recent achievements and conclusion
   - CGA support
   - IPv6 at Université catholique de Louvain
   - Conclusion

1. Introduction
   - Introduction to Shim6
   - Shim6 : a new layer
   - The REAP exploration protocol

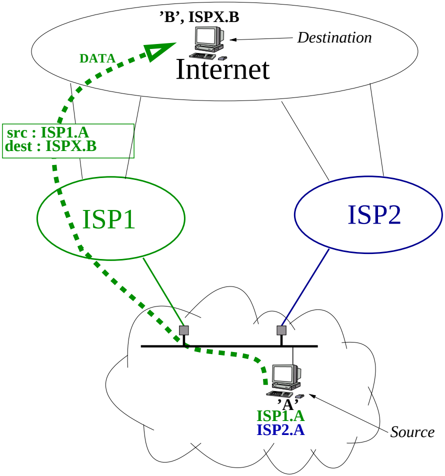
2. The life of an Internet communication...
   - When the TCP SYN is sent...
   - XFRM comes into play
   - Detecting failures : REAP
   - Garbage collection

3. Recent achievements and conclusion
   - CGA support
   - IPv6 at Université catholique de Louvain
   - Conclusion

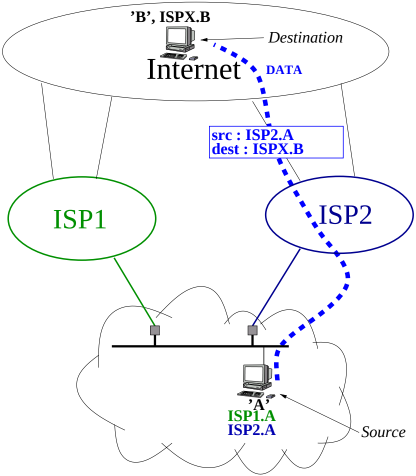
Host-centric multihoming (the context)



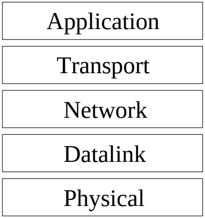
Host-centric multihoming (the context)



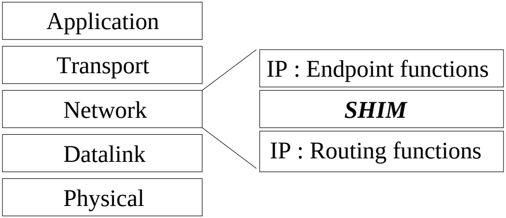
Host-centric multihoming (the context)



Locators vs Identifiers (ULIDs)


Application
Transport
Network
Datalink
Physical

Locators vs Identifiers (ULIDs)



Application

Transport

Network

Datalink

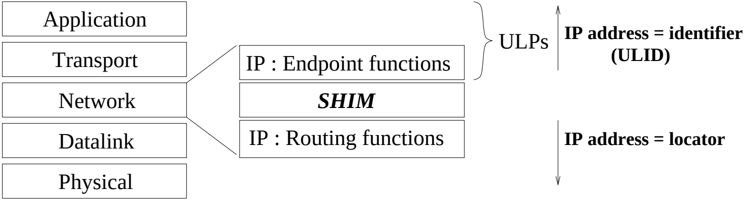
Physical

IP : Endpoint functions

SHIM

IP : Routing functions

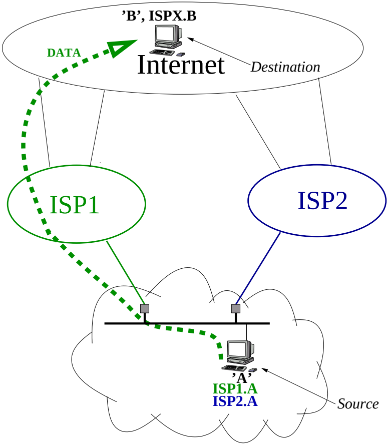
Locators vs Identifiers (ULIDs)



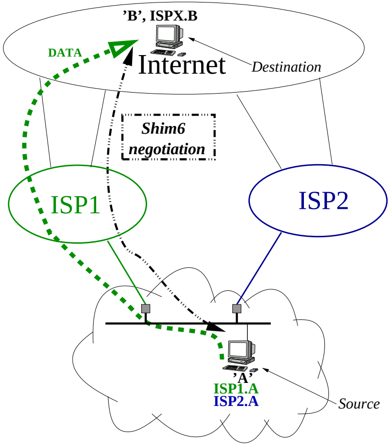
Locators vs Identifiers (ULIDs)

- ULID : Used as the identifier throughout a transport connection.
- locator : IPv6 address used for routing (locating the peer).
- Shim6 performs a mapping between ULIDs and locators, by use of context tags.

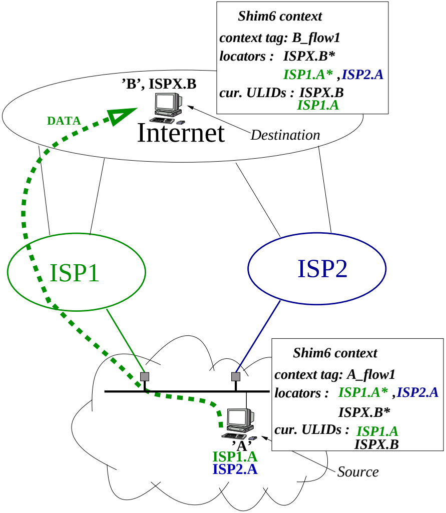
Shim6 operation



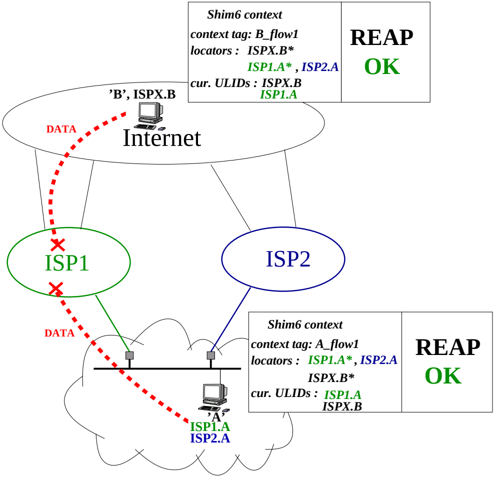
Shim6 operation



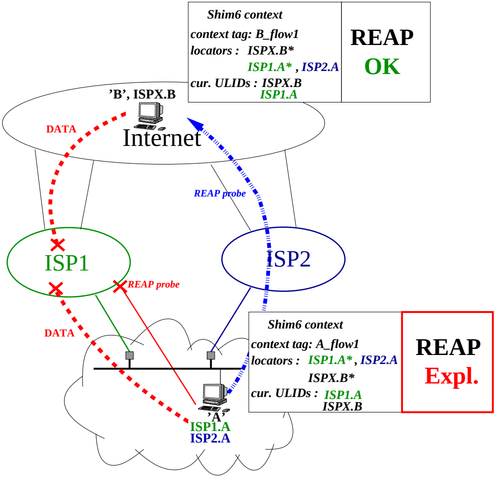
Shim6 operation



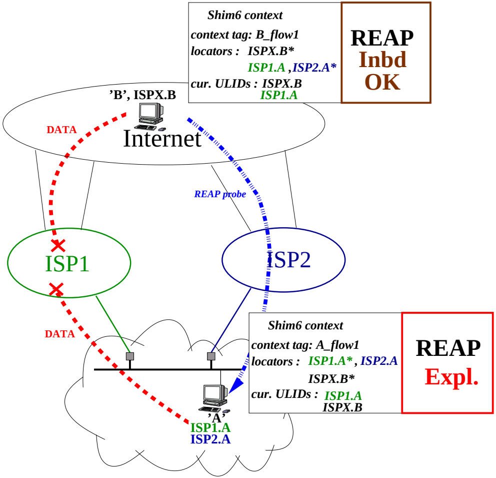
REAP operation



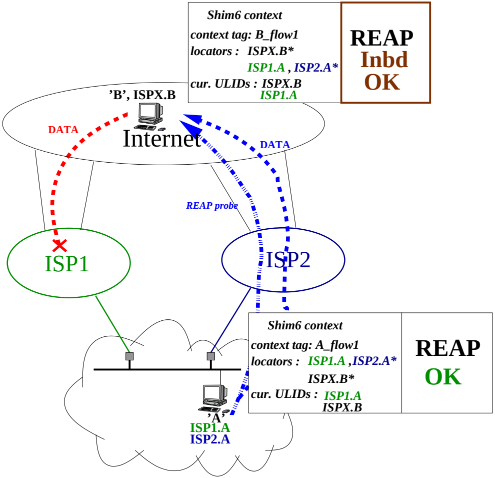
REAP operation



REAP operation



REAP operation



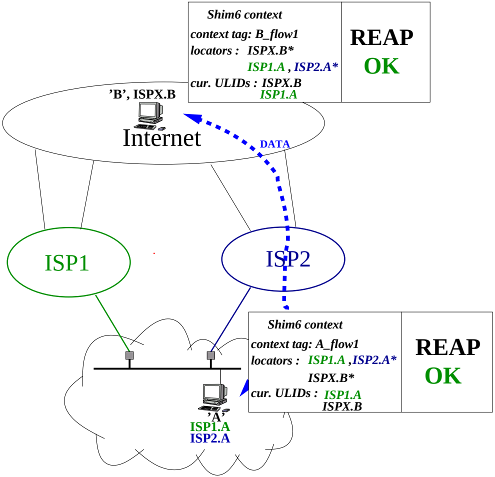
Shim6 context
context tag: B_flow1
*locators : ISPX.B**
*ISP1.A , ISP2.A**
cur. ULIDs : ISPX.B
ISP1.A

REAP
Inbd
OK

Shim6 context
context tag: A_flow1
*locators : ISP1.A ,ISP2.A**
*ISPX.B**
cur. ULIDs : ISP1.A
ISPX.B

REAP
OK

REAP operation



Shim6 context
context tag: B_flow1
*locators : ISPX.B**
*ISP1.A , ISP2.A**
cur. ULIDs : ISPX.B
ISP1.A

REAP
OK

Shim6 context
context tag: A_flow1
*locators : ISP1.A ,ISP2.A**
*ISPX.B**
cur. ULIDs : ISP1.A
ISPX.B

REAP
OK

TCP connection survival

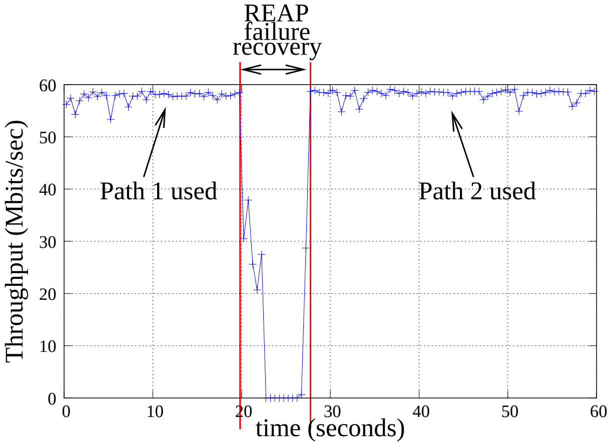


Figure: Evolution of throughput for an iperf TCP session

Introduction : LinShim6

- LinShim6 is developped at INL (UCLouvain) for two years.
- The implementation now supports almost all the Shim6 draft.
- Version 0.5 : Based on IPsec-XFRM framework.
  - Better integration
  - Kernel code minimized.
- CGA support now !

Introduction : LinShim6

- LinShim6 is developped at INL (UCLouvain) for two years.
- The implementation now supports almost all the Shim6 draft.
- Version 0.5 : Based on IPsec-XFRM framework.
    - Better integration
    - Kernel code minimized.
- CGA support now !

Introduction : LinShim6

- LinShim6 is developped at INL (UCLouvain) for two years.
- The implementation now supports almost all the Shim6 draft.
- Version 0.5 : Based on IPsec-XFRM framework.
  - Better integration
  - Kernel code minimized.
- CGA support now !

Introduction : LinShim6

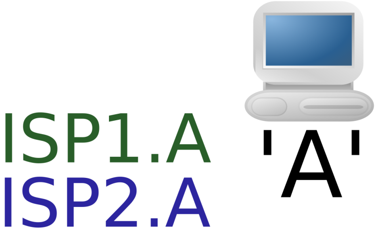
- LinShim6 is developped at INL (UCLouvain) for two years.
- The implementation now supports almost all the Shim6 draft.
- Version 0.5 : Based on IPsec-XFRM framework.
  - Better integration
  - Kernel code minimized.
- CGA support now !

1. Introduction
   - Introduction to Shim6
   - Shim6 : a new layer
   - The REAP exploration protocol

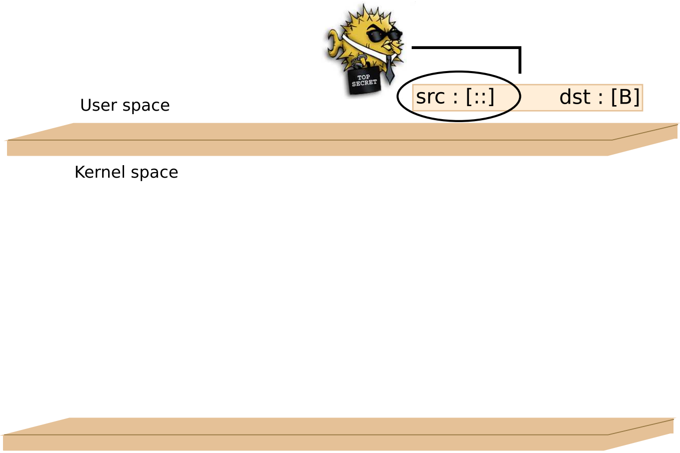
2. The life of an Internet communication...
   - When the TCP SYN is sent...
   - XFRM comes into play
   - Detecting failures : REAP
   - Garbage collection

3. Recent achievements and conclusion
   - CGA support
   - IPv6 at Université catholique de Louvain
   - Conclusion

Starting an SSH exchange

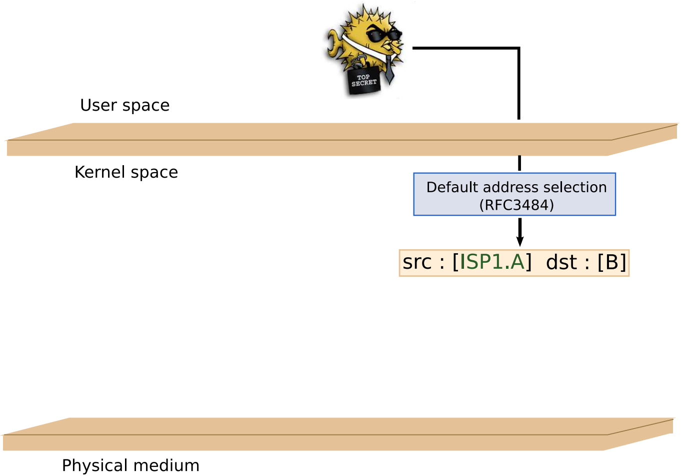


Starting an SSH exchange



Physical medium

Starting an SSH exchange



User space

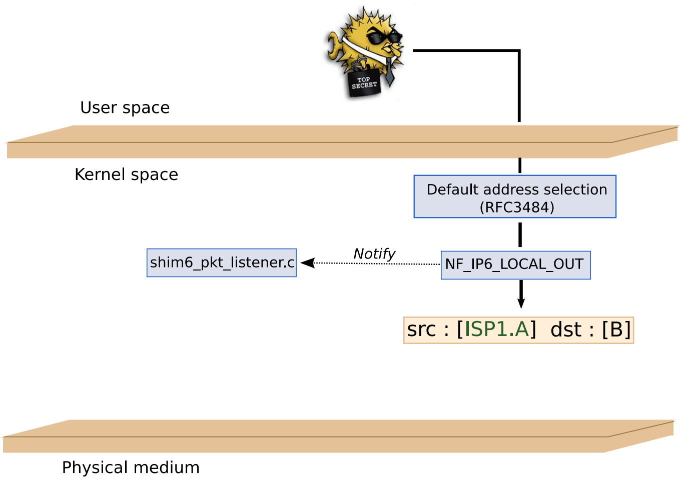
Kernel space

Default address selection (RFC3484)

src : [ISP1.A] dst : [B]

Physical medium

Starting an SSH exchange



User space

Kernel space

Default address selection (RFC3484)

shim6_pkt_listener.c

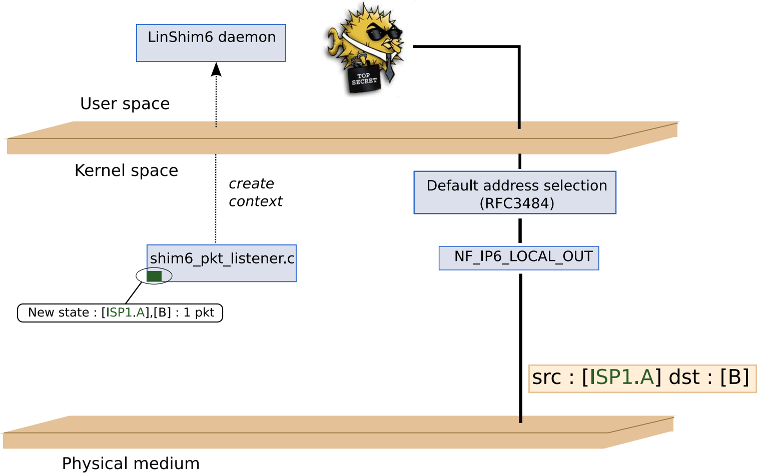
Notify

NF_IP6_LOCAL_OUT

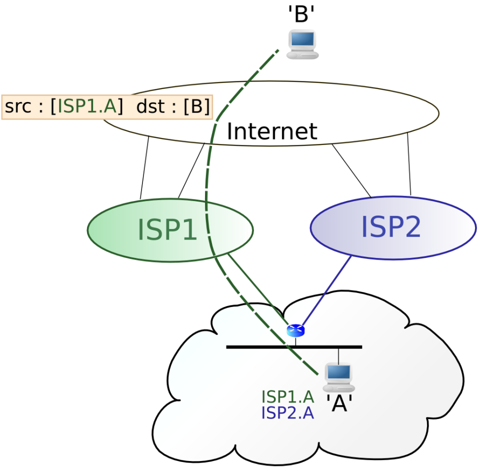
src : [ISP1.A] dst : [B]

Physical medium

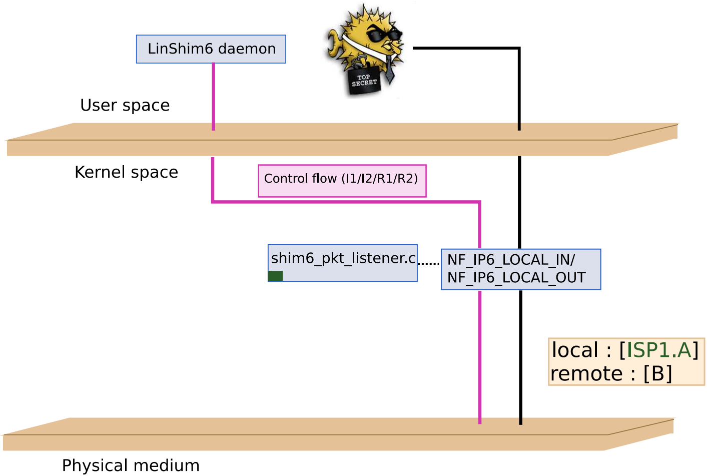
Starting an SSH exchange



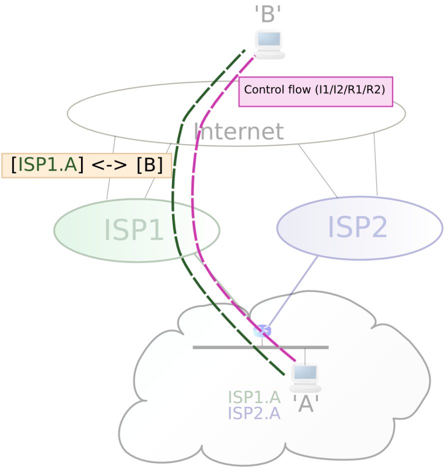
Starting an SSH exchange



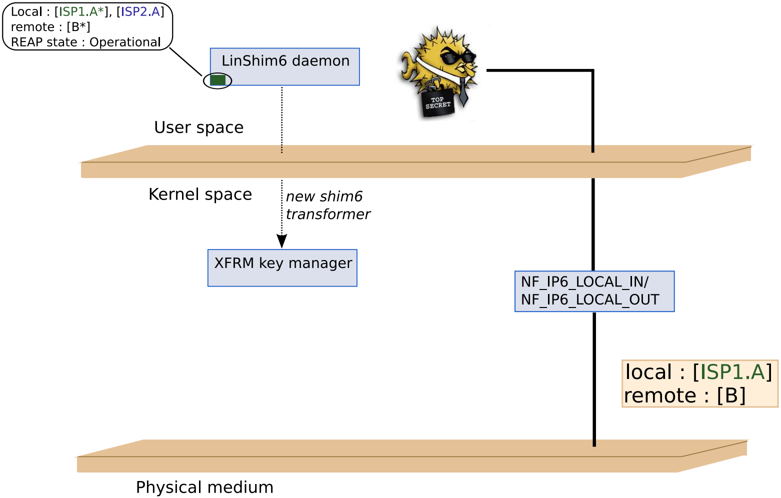
Starting an SSH exchange



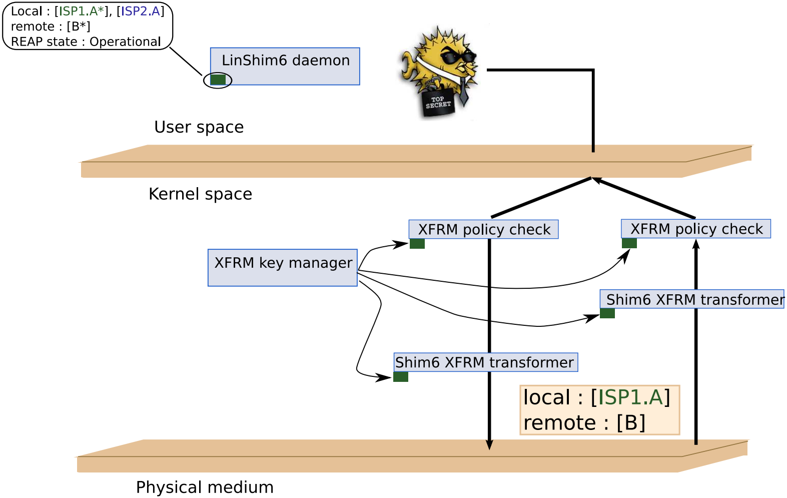
Starting an SSH exchange



The XFRM framework



The XFRM framework



Outbound policy

- Associate a flow with a bundle of transformations.
- Here, policy says :

  if *source* is [ISP1.A] **and** *dest* is [B] **then**
  Modify output path to go through Shim6 Security Association
- The bundle could be AH → ESP → Shim6 (future work).

Inbound policy

- Drop the packet if it does not match any policy.
- Here, policy says :

if not (*source* is [B] **and** *dest* is [ISP1.A]) **then**
Drop packet

Policies vs transformers

- Transformers maintain state (e.g. REAP timers)
- A policy uses generic XFRM code, a transformer has its own code (shim6 address rewriting).
- Shim6 transformers do not necessarily perform translation.
- . . . But they always perform failure detection.

Policies vs transformers

- Transformers maintain state (e.g. REAP timers)
- A policy uses generic XFRM code, a transformer has its own code (shim6 address rewriting).
- Shim6 transformers do not necessarily perform translation.
- . . . But they always perform failure detection.

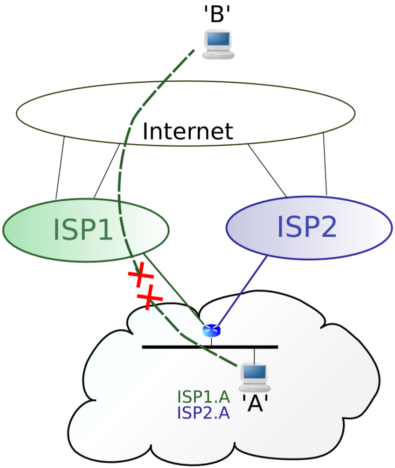
Policies vs transformers

- Transformers maintain state (e.g. REAP timers)
- A policy uses generic XFRM code, a transformer has its own code (shim6 address rewriting).
- Shim6 transformers do not necessarily perform translation.
- . . . But they always perform failure detection.

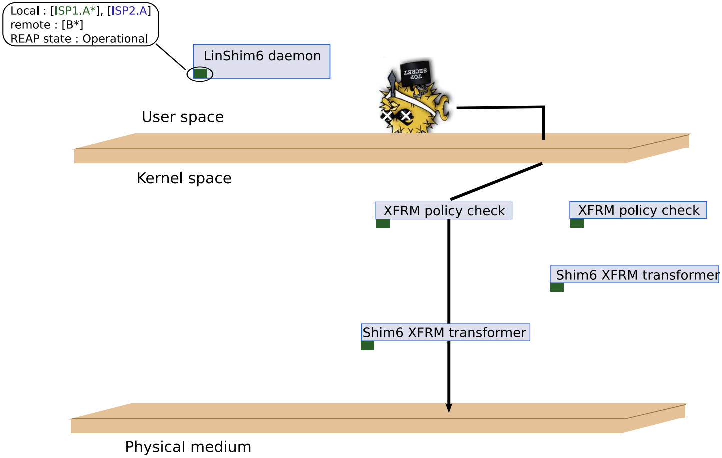
Policies vs transformers

- Transformers maintain state (e.g. REAP timers)
- A policy uses generic XFRM code, a transformer has its own code (shim6 address rewriting).
- Shim6 transformers do not necessarily perform translation.
- . . . But they always perform failure detection.

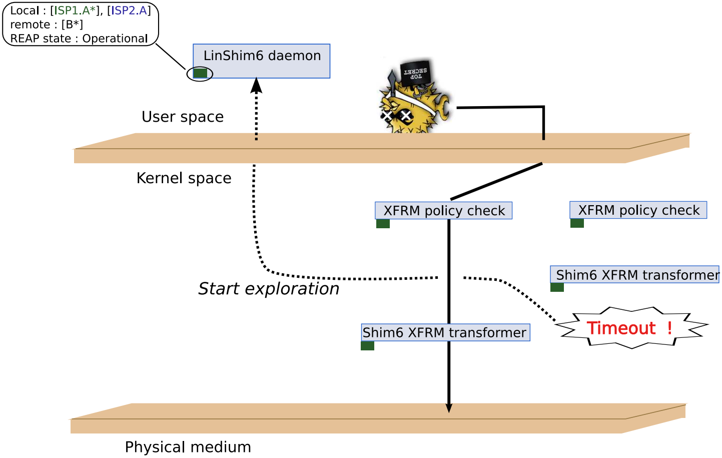
Failure detection



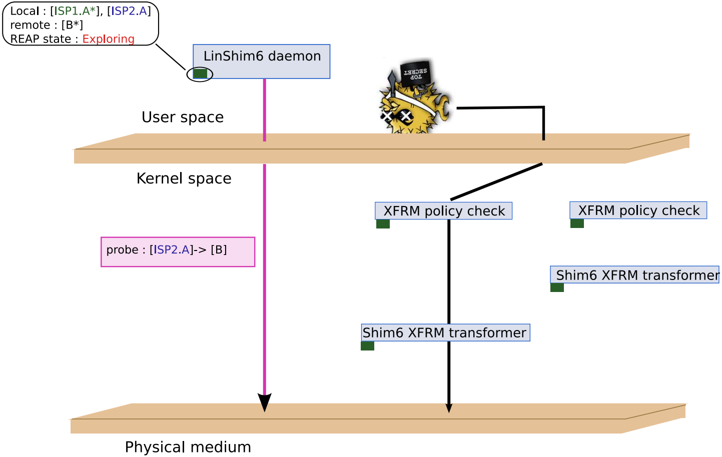
Failure detection



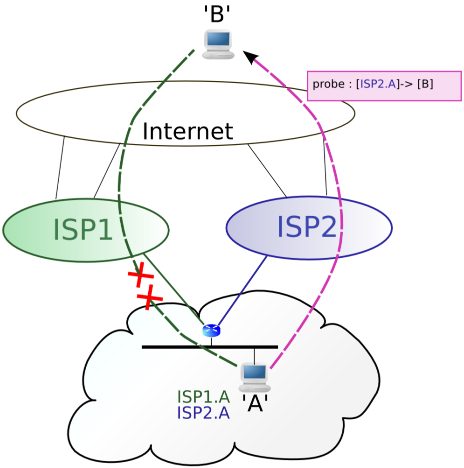
Failure detection



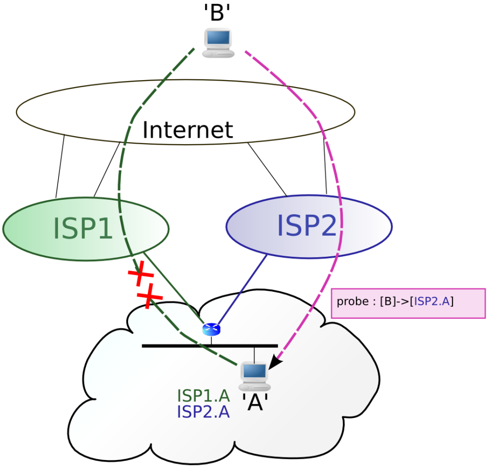
Failure detection



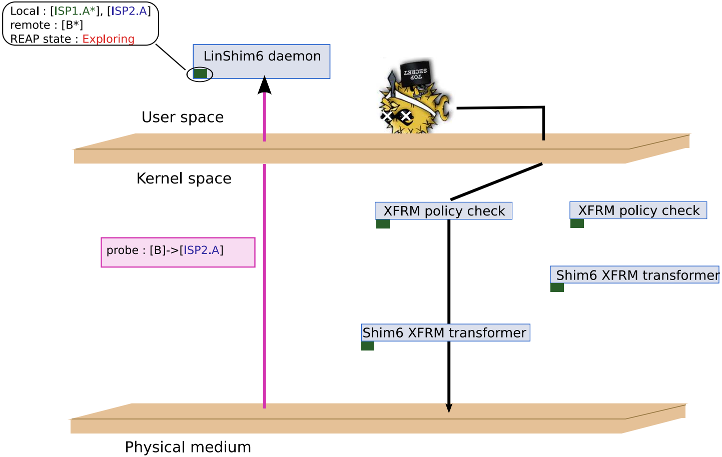
Failure detection



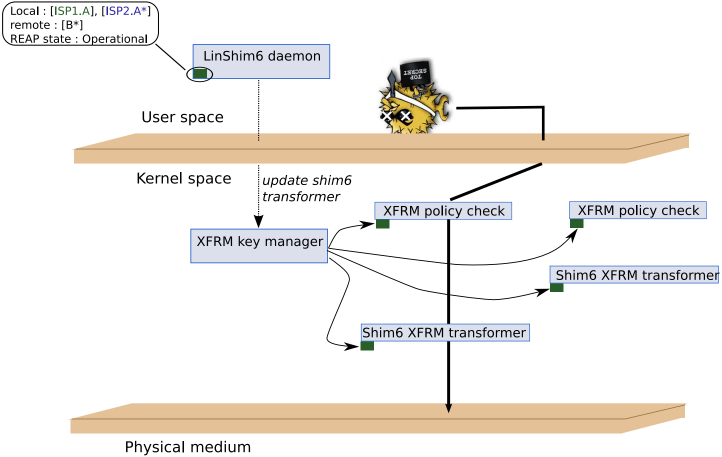
Failure detection



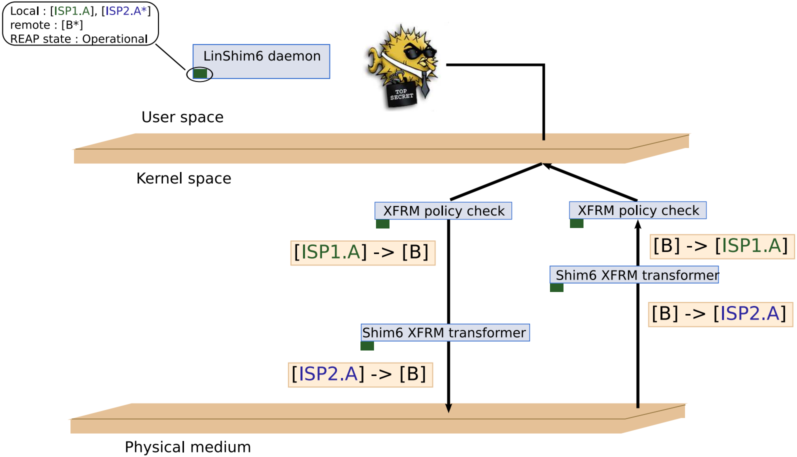
Failure detection



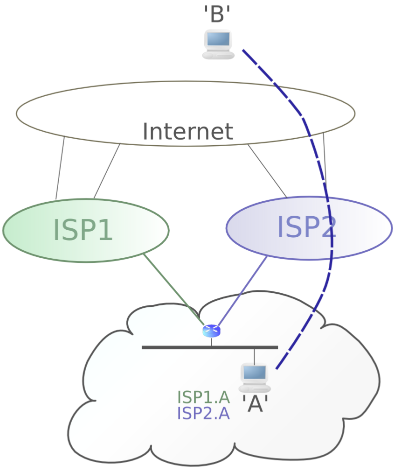
Failure detection



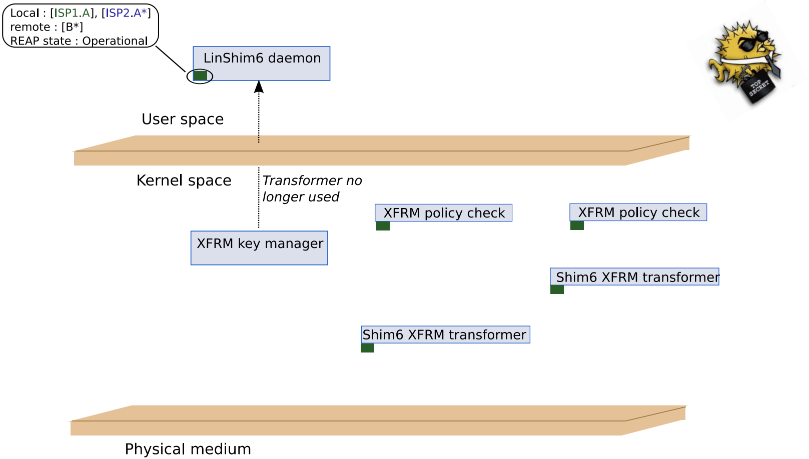
Failure detection



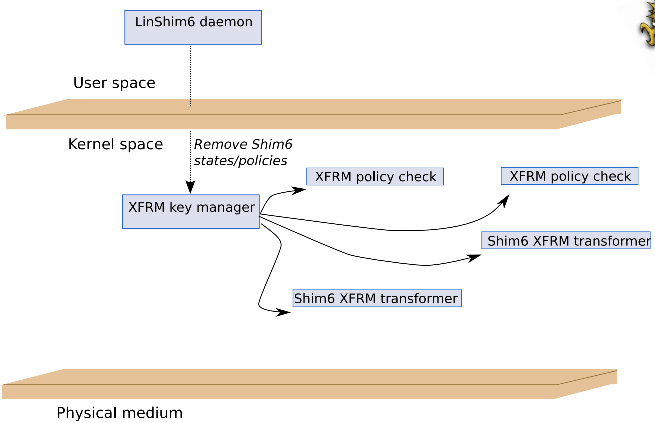
Failure detection



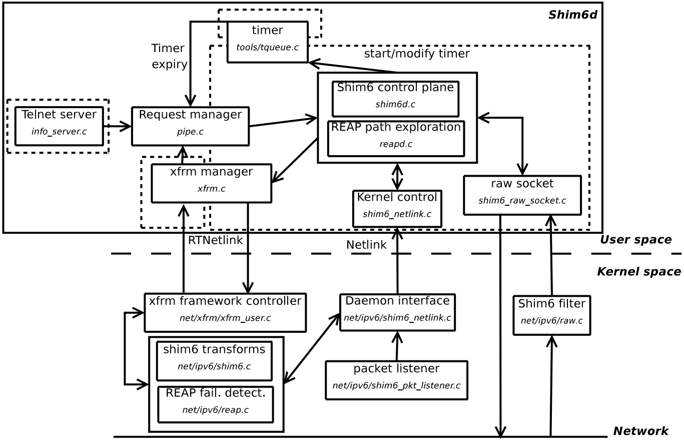
Garbage collecting a Shim6 context



Garbage collecting a Shim6 context



Summarizing the architecture



1 Introduction
- Introduction to Shim6
- Shim6 : a new layer
- The REAP exploration protocol

2 The life of an Internet communication...
- When the TCP SYN is sent...
- XFRM comes into play
- Detecting failures : REAP
- Garbage collection

3 Recent achievements and conclusion
- CGA support
- IPv6 at Université catholique de Louvain
- Conclusion

CGA now supported !

- Implementation is based on DoCoMo SEcure Neighbor Discovery
  - `http://www.docomolabs-usa.com/lab_opensource.htm`
- CGA daemon auto-generates CGA based on RAs
- LinShim6 now only accepts secured addresses.

CGA now supported !

- Implementation is based on DoCoMo SEcure Neighbor Discovery
  - `http://www.docomolabs-usa.com/lab_opensource.htm`
- CGA daemon auto-generates CGA based on RAs
- LinShim6 now only accepts secured addresses.

CGA now supported !

- Implementation is based on DoCoMo SEcure Neighbor Discovery
  - http://www.docomolabs-usa.com/lab_opensource.htm
- CGA daemon auto-generates CGA based on RAs
- LinShim6 now only accepts secured addresses.

IPv6 state at UCL

- Native IPv6 through Belnet.
- New sixxs tunnel to Easynet obtained recently
  - All our department will soon be IPv6-multihomed. (one-two weeks)
  - Only 3 – 4 ms RTT to the tunnel broker.
  - http://www.sixxs.net/
- Shim6 experimental server : Future.

IPv6 state at UCL

- Native IPv6 through Belnet.
- New sixxs tunnel to Easynet obtained recently
  - All our department will soon be IPv6-multihomed. (one-two weeks)
  - Only 3 – 4 ms RTT to the tunnel broker.
  - `http://www.sixxs.net/`
- Shim6 experimental server : Future.

IPv6 state at UCL

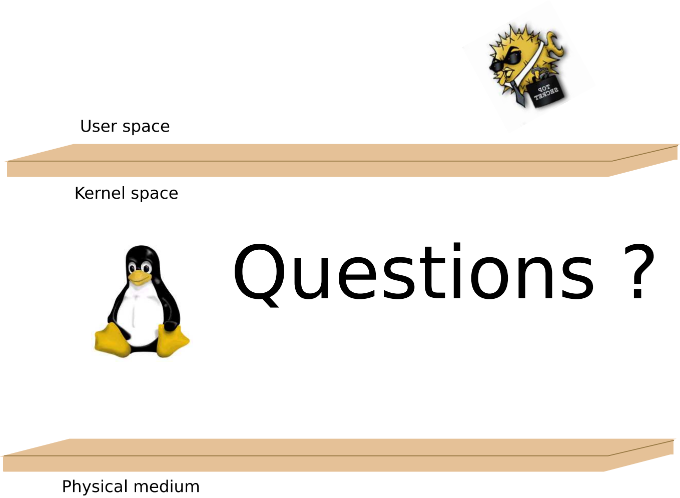
- Native IPv6 through Belnet.
- New sixxs tunnel to Easynet obtained recently
  - All our department will soon be IPv6-multihomed. (one-two weeks)
  - Only 3 – 4 ms RTT to the tunnel broker.
  - `http://www.sixxs.net/`
- Shim6 experimental server : Future.

Conclusion

- The XFRM framework has several advantages :
  - Less Shim6 specific code
  - Easier interoperation with IPsec
- kernel parts : Only those critical for efficiency.
- CGA support.
- Future/ongoing work :
  - Provide a public environment for Shim6 experiments/measurements.
  - Stabilize the implementation (need for feedback !), make it user friendlier.
  - HBA support.

Acknowledgements

- USAGI team : Shinta Sugimoto, Masahide Nakamura
- DoCoMo : SEND implementation
- LinShim6 users : John Ronan, Lu Junxiu, ENST-Bretagne
- Matthijs Mekking : Wireshark patch for Shim6.
- And several others, thanks for fruitful discussions !



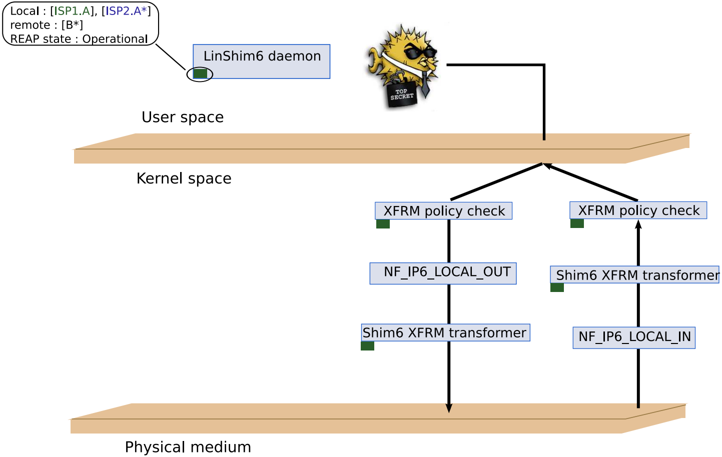
User space

Kernel space

Questions ?

Physical medium

Shim6 XFRM transformers vs Netfilter


Local : [ISP1.A], [ISP2.A*]
remote : [B*]
REAP state : Operational
LinShim6 daemon
TOP SECRET
User space
Kernel space
XFRM policy check
NF_IP6_LOCAL_OUT
Shim6 XFRM transformer
XFRM policy check
Shim6 XFRM transformer
NF_IP6_LOCAL_IN
Physical medium