

SWIN
BUR
NE

SWINBURNE
UNIVERSITY OF
TECHNOLOGY

MPTCP for FreeBSD

lastewart@swin.edu.au

Centre for Advanced Internet Architectures (CAIA)
Swinburne University of Technology





Who Is This Guy?

- BEng (Telecomms and Internet Technologies) / BSci (Comp Sci and Software Eng) (2001-2006)
- Centre for Advanced Internet Architectures, Swinburne University (2003-2013)
 - Research assistant/engineer during/after studies
 - PhD candidate, transport protocol dynamics (2007-)
 - FreeBSD user since 2003, developer since 2008
 - <http://caia.swin.edu.au/cv/lstewart>
- Netflix OpenConnect team (2013-)



What Do We Get?

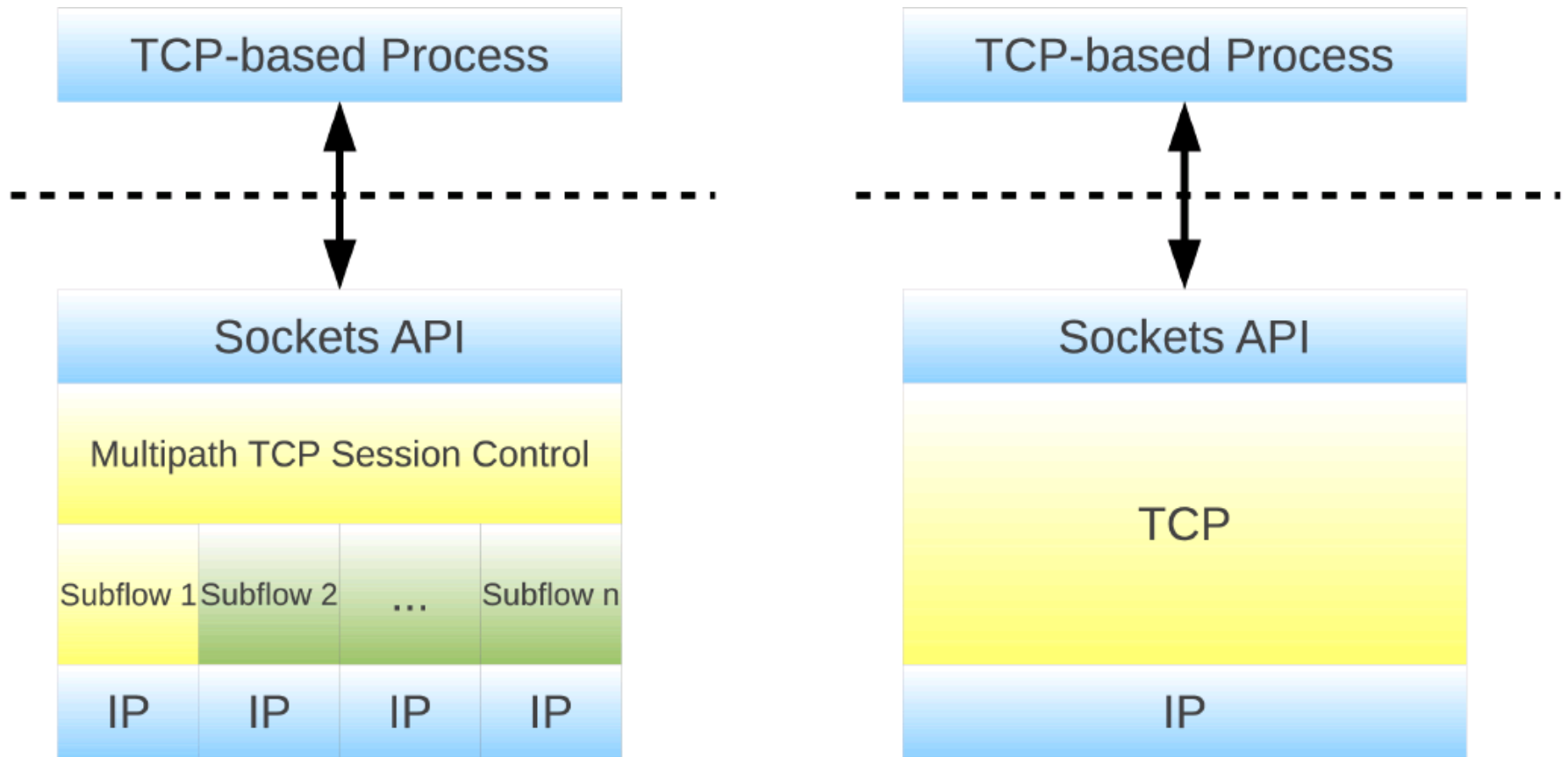
- IP address-based multipath
- Connection redundancy
- Data striping
- “Make-after-break” connections
- Transparent to legacy applications
- Coupled congestion control
- 64bit sequence space
- Significantly increased complexity

Architecture: Integration With TCP



- Shim tightly coupled with TCP code
- Migrate session management into shim
- Tweak control data structure relationships
- RX side
 - Merge TCP reassembly + in-order delivery queue
 - Defer data-level reassembly to user context
- TX side
 - Map chunks of socket buffer to subflows

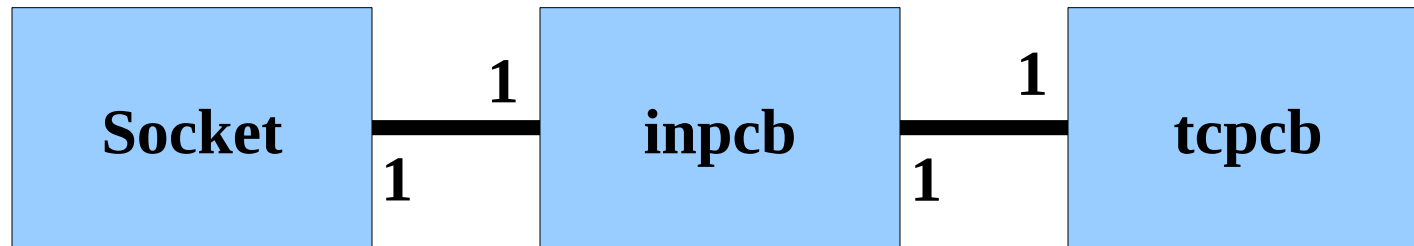
Architecture: Logical Layering



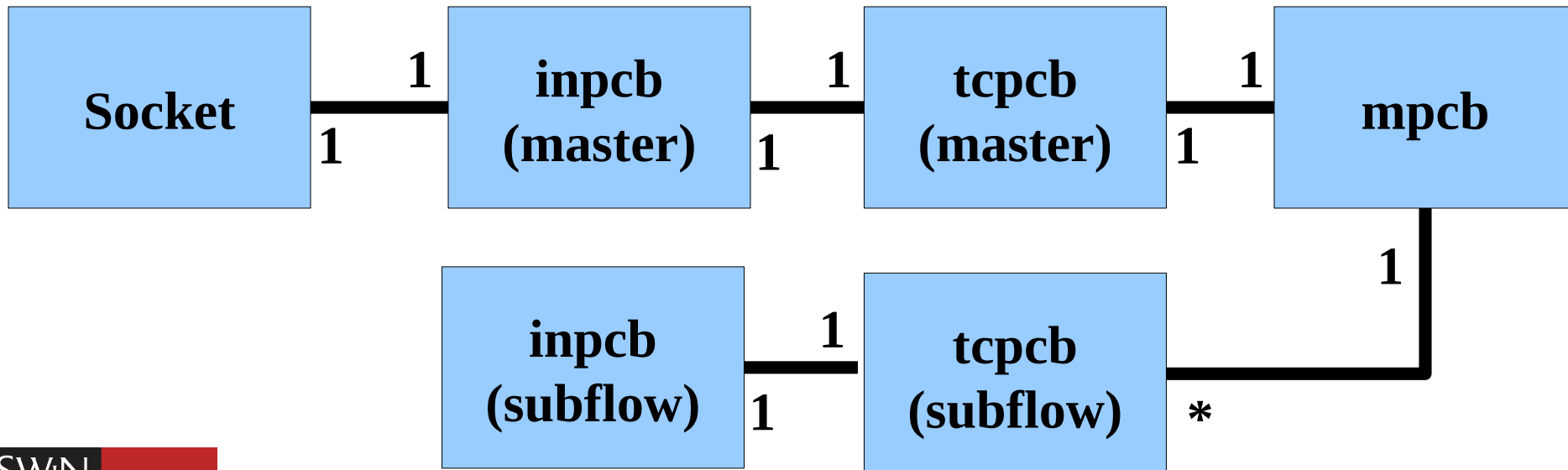
Architecture: Control Data Structures



Before:



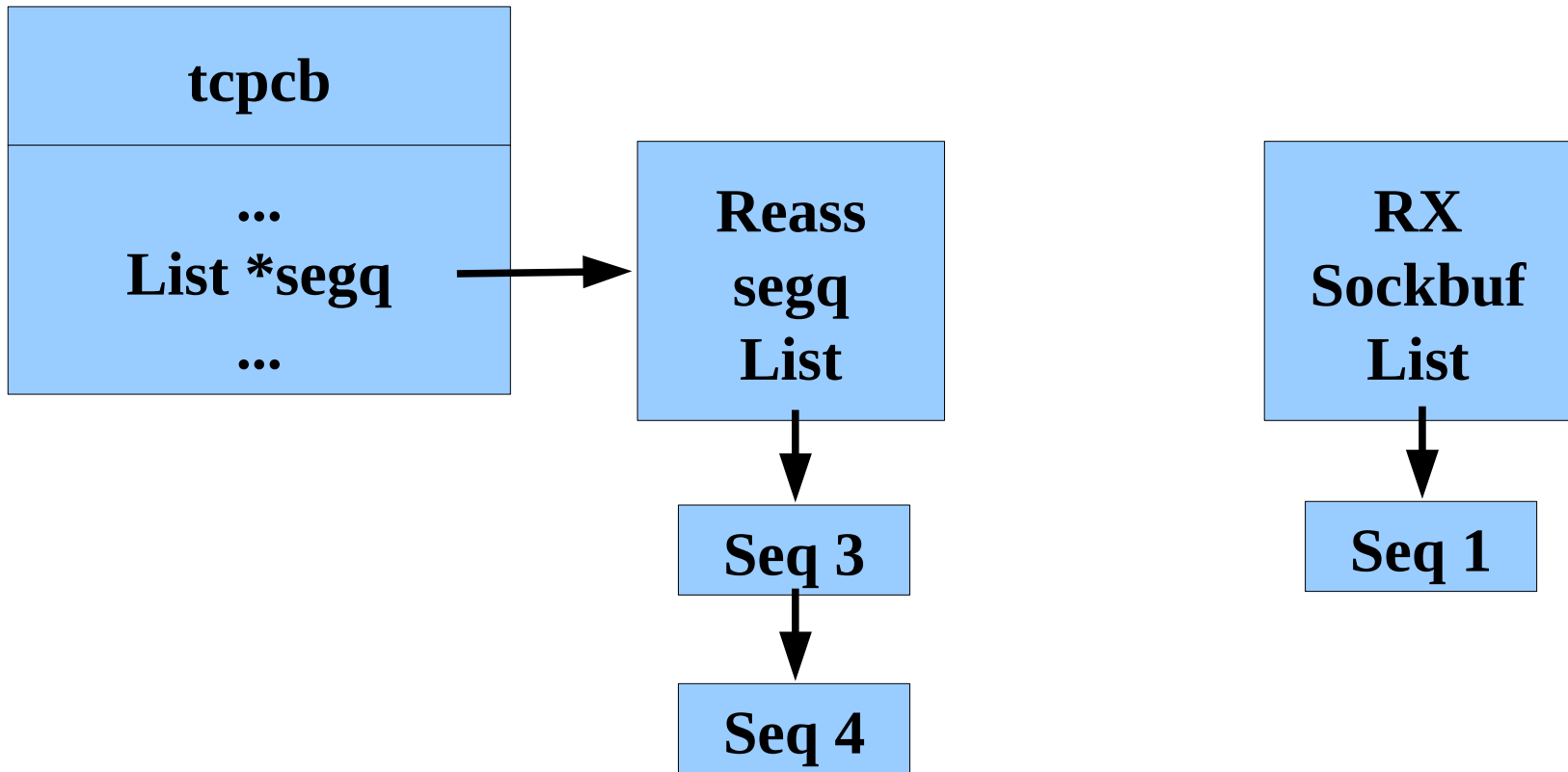
• After:



Architecture: RX Data Structures



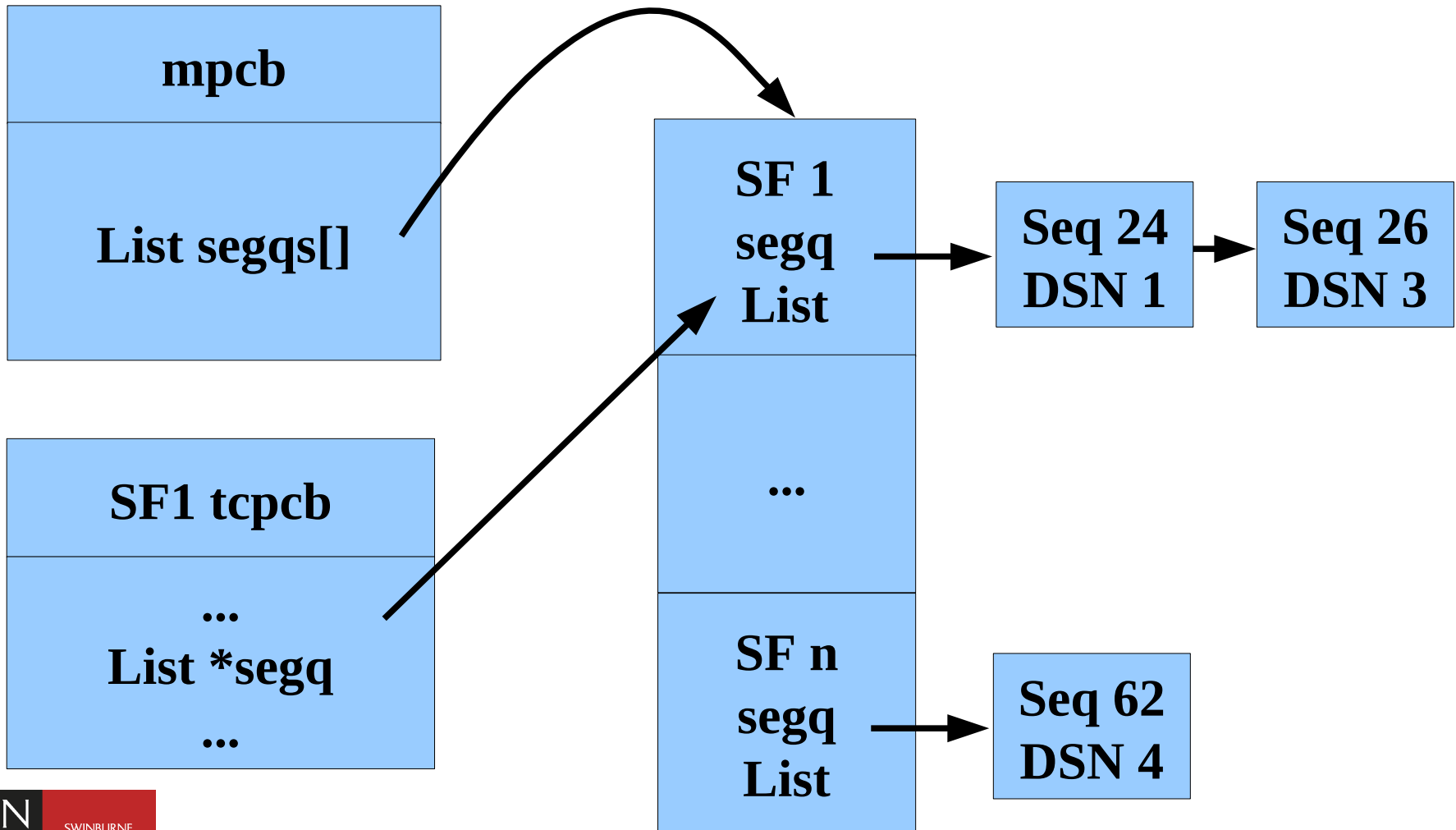
Before:



Architecture: RX Data Structures



After:

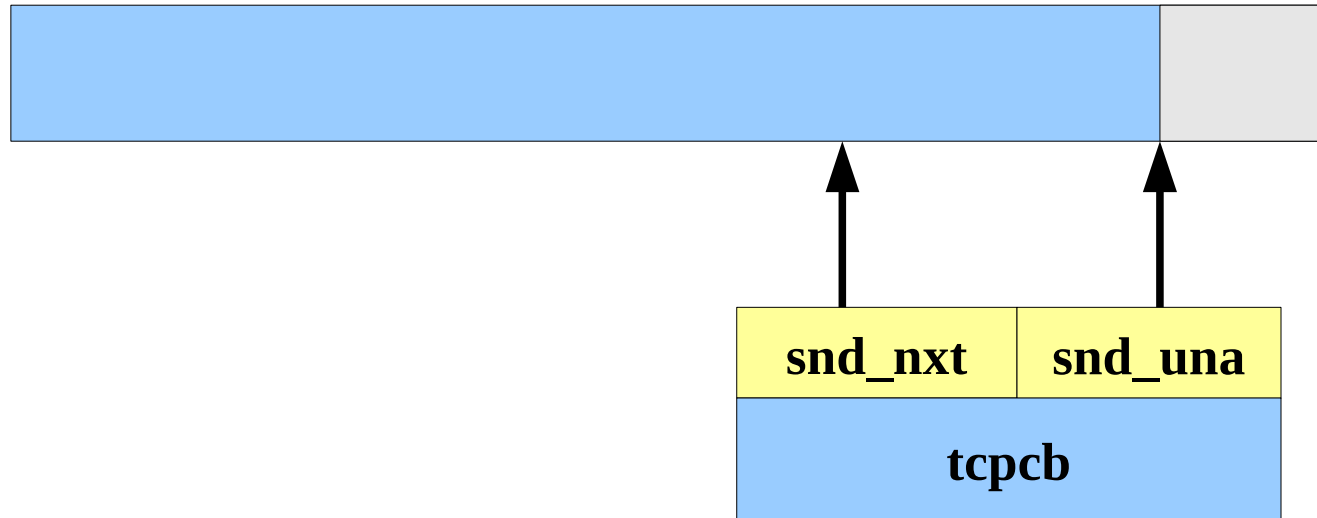


Architecture: TX Data Structures



Before:

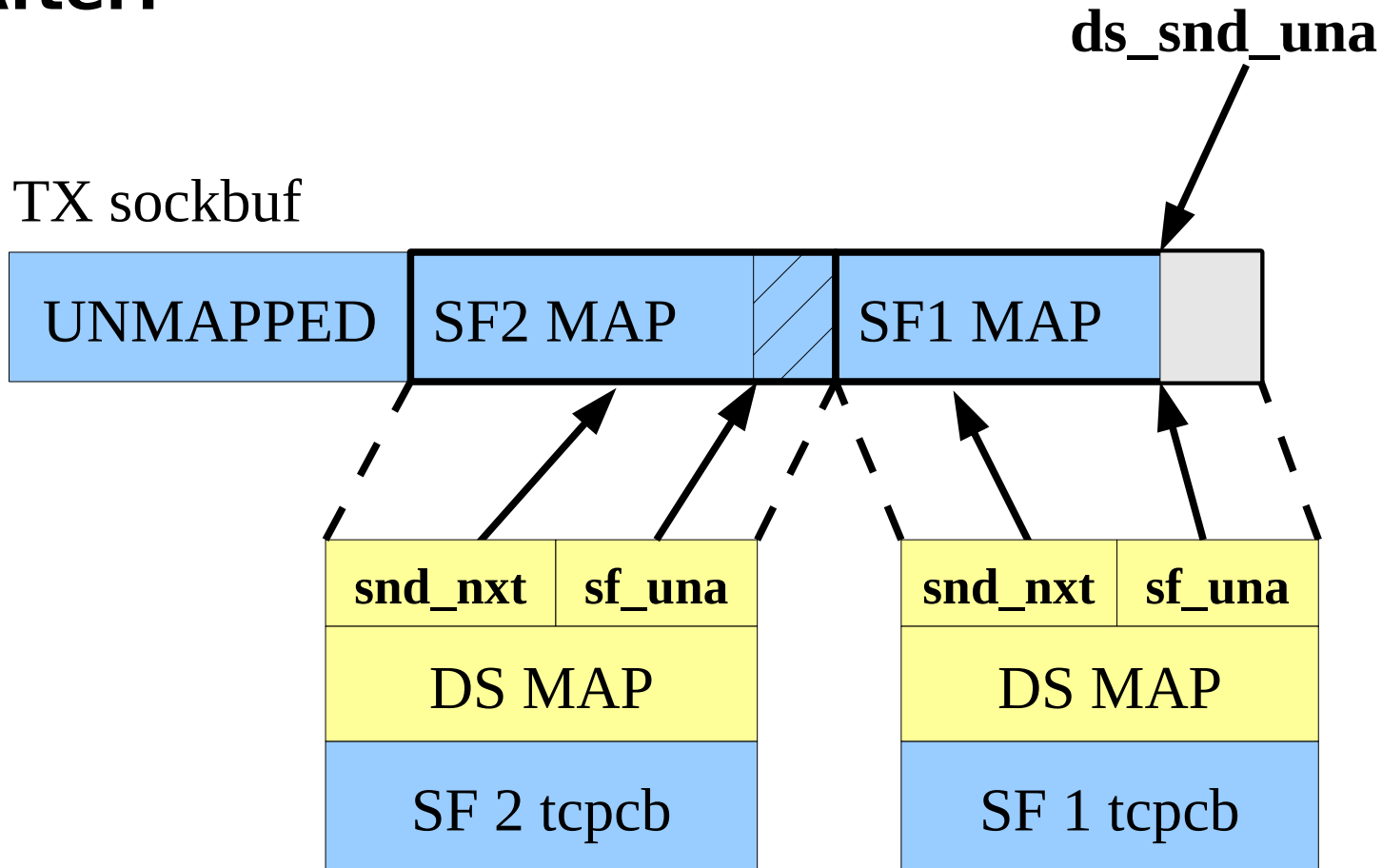
TX sockbuf



Architecture: TX Data Structures



After:



Architecture: SMP

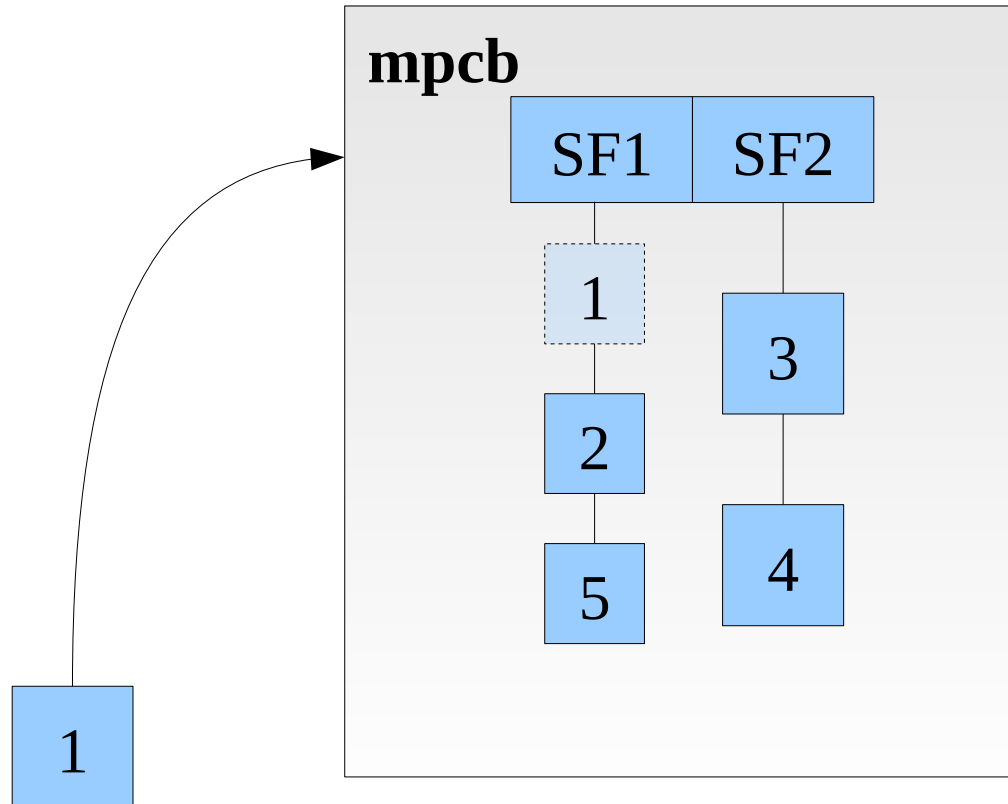


- Reader-writer locks, sensible data structures & access patterns to minimise lock contention
- TX
 - Sub-flows rlock sockbuf to send mapped data
 - MP shim wlocks sockbuf when allocating new map or freeing ACKed data
 - User context wlocks sockbuf to write()
- RX
 - Sub-flows rlock seg queues array to enqueue
 - User context wlocks seg queues array to reassemble

Architecture: RX Data Delivery



Insert into segment list

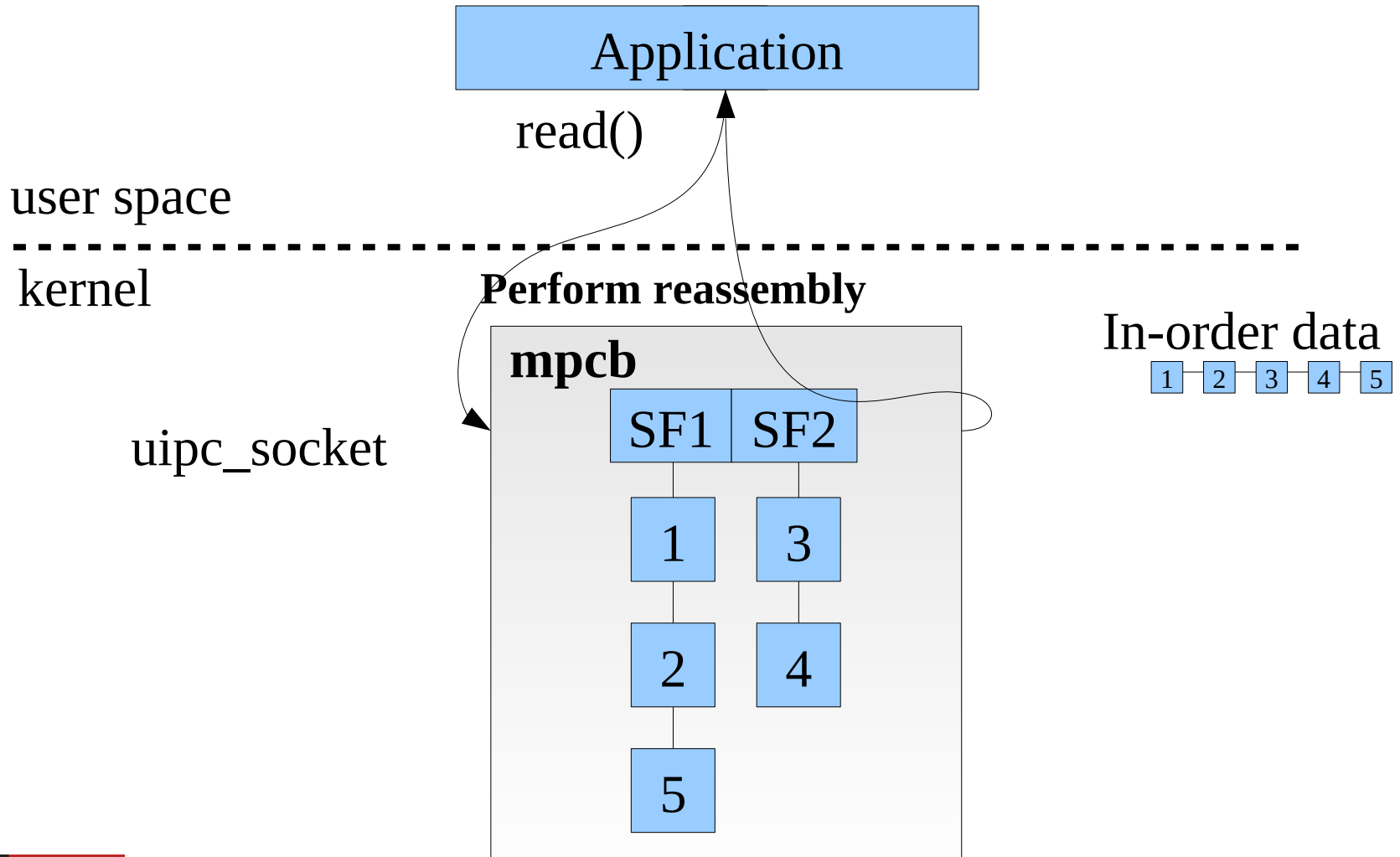


Segment fills hole.
Call 'sorwakeup' to
wake process

Segment arrives
on subflow 1

Schedule subflow
and data-level ACKs

Architecture: RX Data Delivery



News



- IETF progressed drafts to experimental RFCs
- iOS7 has MPTCP
- v0.4 FreeBSD patch due out “any day now”™
 - Fixes numerous stability problems
 - We will push into FreeBSD project branch after release
- Active areas of research at CAIA:
 - Per-subflow congestion control
 - Subflow packet scheduling
 - Vehicle-to-infrastructure with MPTCP

Acknowledgements



CISCO™



Questions?



<http://caia.swin.edu.au/urp/newtcp/mptcp/>