

PROTOKOL ZA UPRAVLJANJE PRENOSOM TOKA PODATAKA

(SCTP – Stream Control
Transmission Protocol)

Uvod, istorijat

- ◆ Slično kao TCP i UDP, SCTP je transportni protokol za komunikaciju s-kraja-na-kraj (eng. end-to-end).
- ◆ Originalno je projektovan unutar IETF radne grupe SIGTRAN (Signaling Transport).
- ◆ Cilj je bio da se otklone nedostaci TCP-ja koji se tiču prenosa telefonske signalizacije preko IP mreža.
- ◆ Konkretni problem je bio prenos signalizacije SS7 preko Interneta.
- ◆ Kasnije je SCTP evoluirao u opštenamenski protokol sa realizacijama u jezgrima raznih operativnih sistema.
- ◆ Osnovna preporuka je RFC 4960.

Uvod, namena

- ◆ Slično kao TCP, SCTP obezbeđuje kom. vezu na nivou 4, koji je:
 - zasnovan na uspostavi veze (eng. connection-oriented),
 - sa pouzdanim prenosom podataka (tj. bez gubitaka),
 - potpuno dvosmeran (eng. full-duplex), i
 - sa kontrolom toka i preopterećenja (eng. flow control, congestion control).
- ◆ Za razliku od TCP i UDP, SCTP nudi nove opcije isporuke podataka.
- ◆ Ove opcije više odgovaraju raznovrsnim potrebama modernih aplikacija.

Uspostava veze i prenos podataka

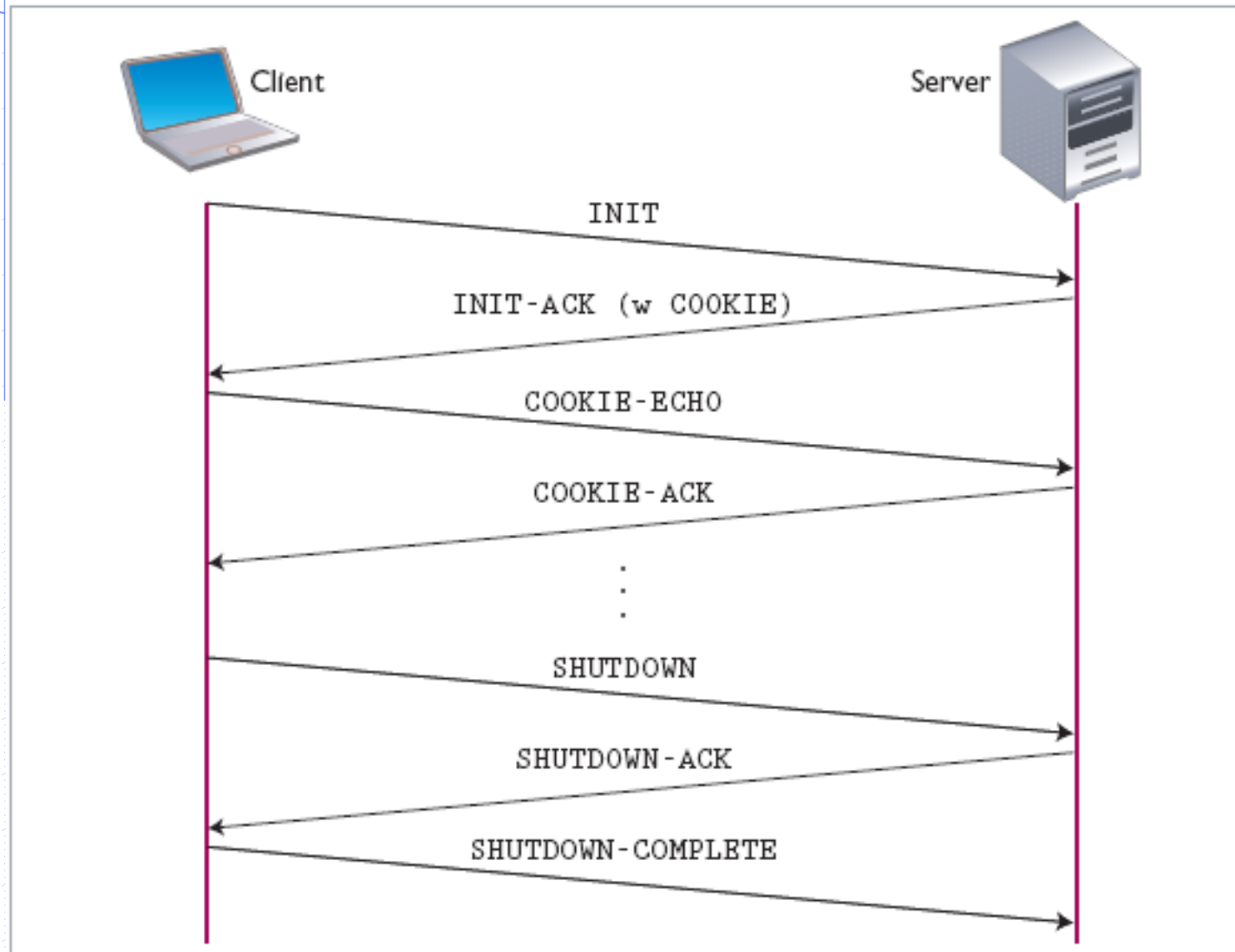
- ◆ Slično kao u TCP, SCTP krajnje tačke prvo pregovaraju i uspostavljaju SCTP vezu (eng. *SCTP association*).
- ◆ Nakon toga sledi faza prenosa podataka.
- ◆ Nakon završetka faze prenosa podataka, krajnje tačke raskidaju transportnu vezu (SCTP association).

Uspostava veze (1/4)

- ◆ Za razliku od TCP, uspostava SCTP veze je zasnovana na razmeni 4 poruke (four-way handshake) sa mehanizmom provere autentičnosti korisnika (cookie mechanism).
- ◆ Mehanizam provere autentičnosti korisnika štiti server (listener) od napada čiji cilj je sprečavanje posluživanja (denial-of-service).
- ◆ Napadi ovog tipa su zasnovani na lažnom predstavljanju (spoofed connection setup requests).
- ◆ Ovi napadi u kontekstu TCP-ja se nazivaju TCP-SYN napadi (TCP-SYN je TCP poruka za uspostavu veze).

Uspostava veze (2/4):

MSC dijagram za uspostavu i raskid veze



Uspostava veze (3/4)

- ◆ Kao što se vidi na slici, nakon što SCTP server primi zahtev za uspostavu veze (SCTP INIT chunk), on ne dodeljuje resurse potrebne za vezu automatski.
- ◆ Umesto toga, server vraća potvrdu (SCTP INIT-ACK chunk) sa informacijom potrebnom za proveru korisnika (cookie), koja sadrži:
 - sve potrebne detalje koji su potrebni da bi se identifikovao i obradio zahtev za vezom (association),
 - vremenski period važenja informacije za proveru korisnika (cookie's life span),
 - Potpis koji je potreban za proveru integriteta i autentičnosti informacije za proveru korisnika (cookie-a)

Uspostava veze (4/4)

- ◆ Legitimni korisnik vraća nazad serveru informaciju za proveru korisnika (SCTP COOKIE_ECHO).
- ◆ Server zatim dodeljuje potrebne resurse i uspostavlja vezu sa korisnikom.
- ◆ S obzirom da se podaci mogu prenositi već nakon razmene treće poruke (SCTP COOKIE_ECHO), početno kašnjenje pre prenosa podataka nije već nego kod TCP.
- ◆ Naime, u TCP uspostava veza se obavlja razmenom tri poruke (three-way handshake).

Prenos podataka putem poruka (1/2)

- ◆ Slično kao kod UDP, a za razliku od TCP, SCTP očuvava granice aplikacionih poruka (message).
- ◆ Poruke se cele isporučuju prijemnoj aplikaciji.
- ◆ Podrazumevano je da se podaci za jedan sistemski poziv *send* posmatraju kao jedna poruka.
- ◆ S druge strane, aplikacija može koristiti EXPLICIT_EOR opciju SCTP utičnice (socket) radi upisa jedne poruke preko više *send* poziva.

Prenos podataka putem poruka (2/2)

- ◆ Da bi ostvario prenos podataka zasnovanih na porukama, SCTP najpre otkriva MTU (Maximum Transmission Unit) za putanju poruka.
- ◆ Dalje koristi algoritam fragmentacije i ponovnog sklapanja poruka.
- ◆ Predajnik deli poruke koje su duže od MTU na fragmente, a prijemnik ih sklapa u poruku.
- ◆ Od aplikacija koje koriste SCTP se ne zahtevaju eksplicitne oznake kraja poruke (delimiters), čime se uprošćava analiza poruka (message parsing).

Proširenje za delimičnu pouzdanost (1/2)

- ◆ U TCP i običnom SCTP garantuje se da će svi poslati podaci biti isporučeni prijemnoj aplikaciji.
- ◆ RFC 3758 opisuje proširenje SCTP za delimičnu pouzdanost PR-SCTP (partial reliability SCTP).
- ◆ Ovo proširenje omogućava prenos podataka između predajnika i prijemnika sa delimičnom pouzdanošću.
- ◆ Aplikacija specificira koliko predajnik treba da bude uporan prilikom slanja poruke prijemniku.
- ◆ Definiše se vreme života svake poruke (message lifetime).

Proširenje za delimičnu pouzdanost (2/2)

- ◆ PR-SCTP predajnik pokušava da pošalje poruku u toku njenog životnog veka.
- ◆ Nakon isteka životnog veka, predajnik odbacuje poruku, bez obzira da li je uspešno preneti i potvrđena ili ne.
- ◆ Ova vremenski ograničena pouzdanost je korisna za aplikacije koje regularno generišu nove podatke, čime prethodni podaci postaju zastareli (obsolete).
- ◆ Primer je online igra u kojoj igrač često generiše nove koordinate pozicije ili druge podatke sa kratkotrajnom važnošću.

Isporuka podataka van redosleda

- ◆ Pored isporuke unutar redosleda, SCTP nudi uslugu isporuke podataka van redosleda (unordered data delivery).
- ◆ Poruka označena za ovu vrstu isporuke se isporučuje prijemnoj aplikaciji čim stigne u SCTP prijemnik.
- ◆ Isporučuje se čak i ako je van redosleda u odnosu na prethodno prenete poruke.
- ◆ Za razliku od UDP, SCTP obezbeđuje pouzdan prenos sa kontrolom toka i preopterećenja za ove poruke.
- ◆ RFC 4168 definiše ovakav prenos SIP (Session Initiation Protocol) poruka preko SCTP.

Raskid SCTP veze (1/2)

- ◆ Aplikacije mogu koristiti jednu istu SCTP vezu za prenos podataka pouzdano, delimično pouzdano, u redosledu, i van redosleda.
- ◆ Nakon završetka prenosa podataka, SCTP krajnje tačke raskidaju vezu razmenom 3 poruke (3-way handshake).
- ◆ Vidi donji deo ranije datog MSC dijagrama.
- ◆ Za razliku od TCP, kod SCTP ne postoji raskid veze sa završetkom u poluzatvorenom stanju (half-closed).
- ◆ Kada jedna strana zatvori utičnicu, obe strane moraju raskinuti vezu.

Raskid SCTP veze (2/2)

- ◆ Takođe, SCTP dozvoljava ponovnu upotrebu istih brojeva prolaza (port) nakon raskida veze,
- ◆ Ovo zato što SCTP verifikaciona zastavica omogućava korelaciju dolaznih podataka sa korektnom SCTP vezom.
- ◆ S druge strane, TCP mora da čeka dva puta duže od MSL (Maximum Segment Lifetime) u stanju TIME_WAIT da bi mogao ponovo da upotrebi isti prolaz.

Uporedna analiza SCTP, TCP i UDP

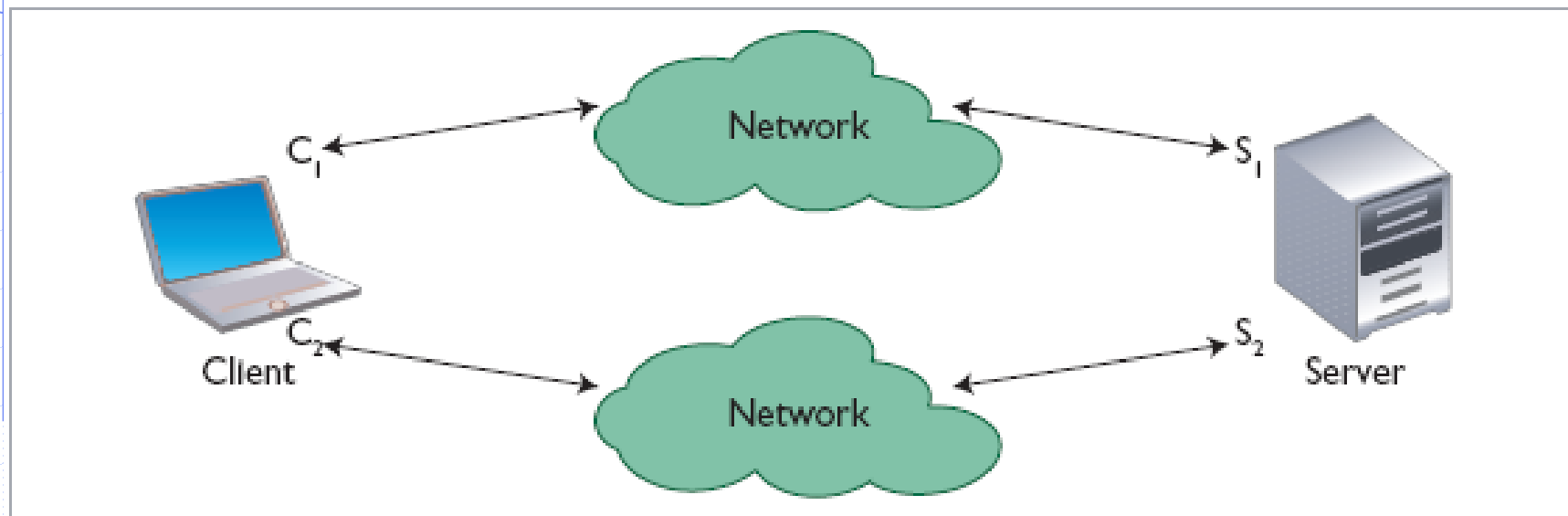
Poređenje SCTP, TCP i UDP

Usluga/Svojstvo	SCTP	TCP	UDP
Komunikacija zasnovana na uspostavi veze	da	da	ne
Prenos podataka zasnovan na porukama	da	ne	da
Pouzdan prenos podataka	da	da	ne
Delimično pouzdan prenos podataka	da	ne	ne
Isporuka podataka unutar redosleda	da	da	ne
Isporuka podataka van redosleda	da	ne	da
Upravljanje tokom i preopterećenjem	da	da	ne
Zaštita od SYN napada putem lažnog predstavljanja	da	ne	-
Dozvoljava poluzatvorene veze	ne	da	-
Više tokova podataka (multistreaming)	da	ne	ne
Više krajnjih IP adresa (multihomig)	da	ne	ne
Dinamičko rekonfigurisanje adresa	da	ne	ne

Krajnje tačke sa više IP adresa (multihoming) (1/3)

- ◆ Dodela višestrukih IP adresa umreženim mašinama i uređajima tehnički je moguća i ekonomski izvodljiva.
- ◆ Korišćenje više mrežnih sprega obezbeđuje otpornost na pojavu mrežnih grešaka s-kraja-na-kraj (end-to-end network fault tolerance).
- ◆ SCTP podržava ovu tehniku na transportnom nivou.
- ◆ Za razliku od TCP i UDP, SCTP veza može da spaja krajnje tačke koje imaju više IP adresa.
- ◆ Na sledećoj slici bi bile moguće 4 različite TCP veze između korisnika (klijenta) i servera:
(C1,S1), (C1,S2), (C2,S1), (C2,S2)

Krajnje tačke sa više IP adresa (multihoming) (2/3)



- ◆ SCTP na transportnom nivou podržava dodelu više IP adresa krajnjim tačkama. Na ovoj slici SCTP spaja dve krajnje tačke sa po dve IP adrese u svakoj tački.

Krajnje tačke sa više IP adresa (multihoming) (3/3)

- ◆ Umesto toga, jedna SCTP veza može uključiti više IP adresa u svakoj krajnjoj tački.
- ◆ U primeru sa slike to bi bilo ($\{C1, C2\}$, $\{S1, S2\}$).
- ◆ SCTP predajnik bira jednu odredišnu adresu kao primarnu, i na nju šalje sve nove podatke.
- ◆ Predajnik stalno nadzire dostupnost svih odredišta.
- ◆ Kad dođe do greške nedostupnosti primarnog odredišta, predajnik dinamički određuje alternativno odredište (*failover*).
- ◆ Alternativnom odredištu šalje podatke sve dok primarno ponovo ne postane dostupno.
- ◆ Tako SCTP obezbeđuje otpornost na pojavu fizičkih otkaza u mreži.

Dinamičko rekonfigurisanje adresa

- ◆ Mnogi moderni računari dozvoljavaju dinamičko uključivanje i isključivanje mrežnih kartica (hot-pluggable interfaces).
- ◆ RFC 5061 proširuje adresnu šemu i omogućava rekonfigurisanje adresa nivoa 3 u radu (on-the-fly).
- ◆ Pomoću DAR (Dynamic Address Reconfiguration), krajnja tačka može dinamički da dodaje i briše IP adrese.
- ◆ Može i da menja primarno odredište u tekućoj vezi.
- ◆ Pored bržeg oporavka od otkaza (failover), omogućava efikasnu primopredaju mobilnih korisnika (hand-off).

Konkurentan prenos podataka po više putanja (Concurrent Multipath Transfer)

- ◆ Krajnje tačke sa više IP adresa mogu se istovremeno povezati preko više putanja s-kraja-na-kraj.
- ◆ Ovo se radi da bi se povećala otpornost na otkaze, npr. kao na prethodnoj slici.
- ◆ Recimo, korisnici mogu istovremeno biti povezani preko telefonske/širokpojasne ili više bežičnih tehnologija, kao što se 802.11b i GPRS.
- ◆ CMT koristi postojanje više nezavisnih putanja za istovremen prenos podataka radi povećanja propusnosti (throughput – količina podataka u jedinici vremena).

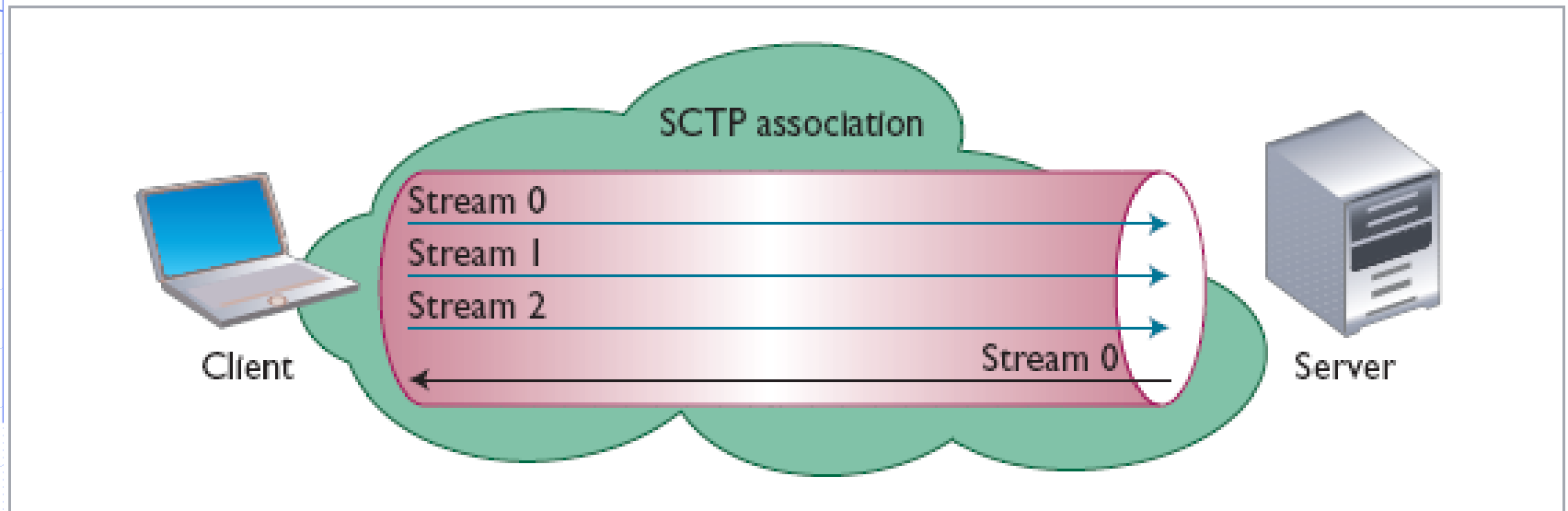
Više tokova (multistreaming) (1/3)

- ◆ SCTP veza se može sastojati od više tokova podataka.
- ◆ Svaki tok je logička, jednosmerna staza unutar veze.
- ◆ Redosled podataka se održava unutar toka, ali ne i između tokova.
- ◆ Svaki tok ima nezavistan prostor redosleda.
- ◆ Prijemnik isporučuje podatke aplikaciji po redosledu toka, bez obzira na relativni redosled podataka koji stižu na drugim tokovima.
- ◆ Tako se nezavisne aplikacione poruke mogu slati po različitim tokovima da bi se očuvala logička izolovanost tokom prenosa i isporuke.

Više tokova (multistreaming) (2/3)

- ◆ Npr. nezavisne HTTP transakcije mogu koristiti nezavisne SCTP tokove radi povećanja performanse Web servera.
- ◆ SCTP veza može imati do 65536 ulaznih ili izlaznih tokova.
- ◆ Krajnje tačke pregovaraju o broju tokova prilikom ili nakon uspostave veze.
- ◆ Npr. klijent može koristiti 3 izlazna toka i 1 ulazni, vidi narednu sliku.
- ◆ Svi tokovi podležu pod zajedničku kontrolu preopterećenja, tj. zadržava se TCP princip ravnopravnosti veza.

Više tokova (multistreaming) (3/3)



- ◆ Primer Sctp veze sa više tokova. Ova veza koristi 3 izlazna toka (od klijenta ka serveru) i 1 ulazni tok.

HTTP preko SCTP (1/3)

- ◆ HTTP zahteva pouzdan transportni protokol.
- ◆ Iako je uglavnom korišćen TCP, HTTP ne zahteva izričito TCP.
- ◆ Nedostatak TCP je što nudi samo jedan tok bajta (okteta) do Web servera.
- ◆ Pod HTTP 1.1 sa perzistencijom i protočnom obradom (pipelining) nezavisni HTTP odgovori se serijalizuju i šalju jedan po jedan preko tog toka okteta.
- ◆ Još gore, TCP garantuje isporuku unutar redosleda po tom toku okteta.

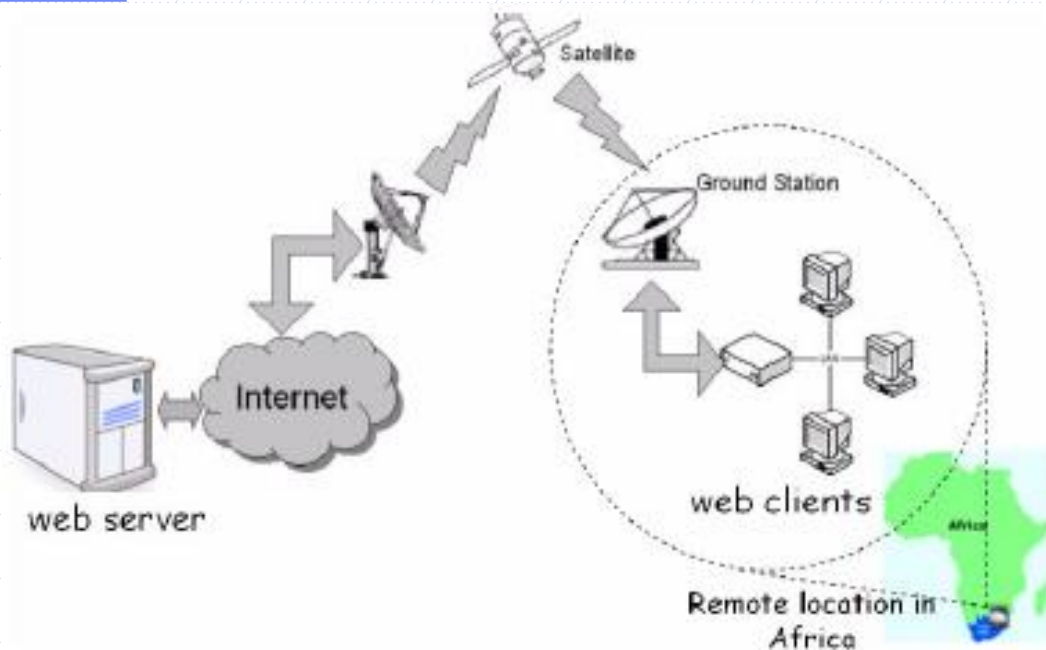
HTTP preko SCTP (2/3)

- ◆ Gubitak jednog dela HTTP odgovora dovodi do kašnjenja u isporuci svih sledećih odgovora.
- ◆ Ovo kašnjenje je jednako vremenu potrebno predajniku da ponovo pošalje izgubljeni deo odgovora.
- ◆ Ovaj problem je poznat pod nazivom HOL (head-of-line) blokiranje.
- ◆ HOL blokiranje potiče od činjenice da TCP ne može logički da razdvoji HTTP odgovore.
- ◆ Ta mogućnost razdvajanja jednostavno ne postoji u TCP mehanizmima transporta i isporuke podataka.

HTTP preko SCTP (3/3)

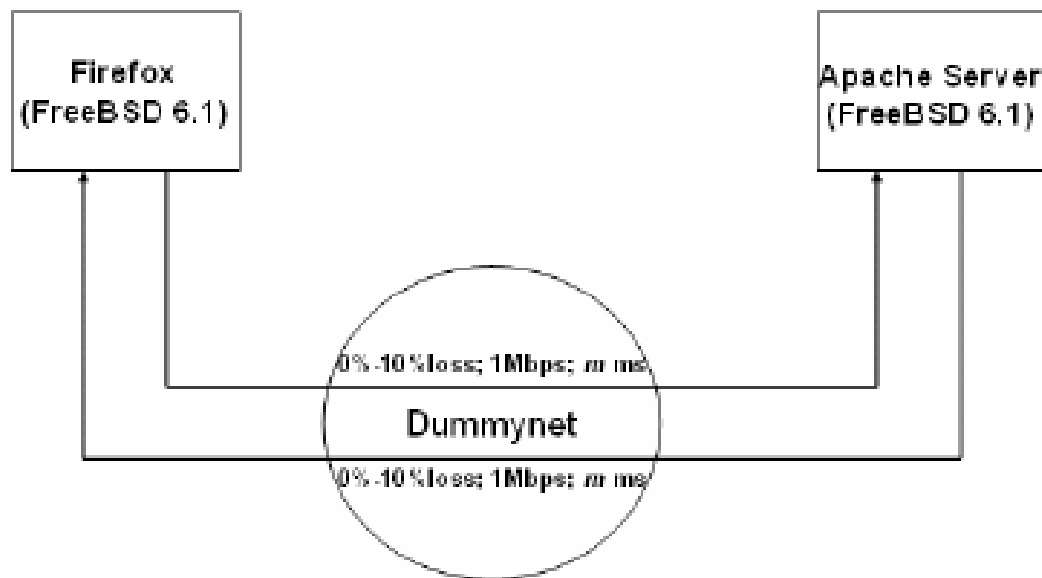
- ◆ U slučaju da se HTTP transakcije prenose po različitim SCTP tokovima, SCTP prijemnik isporučuje HTTP transakcije bez HOL blokiranja između transakcija.
- ◆ Naredna studija slučaja sa velikim kašnjenjem i visokom stopom grešaka u prenosu opravdava ovaj pristup.
- ◆ U studiji se koriste SCTP-sposoban Apache Web server i Firefox Web pretraživač.
- ◆ U toku je identifikovanje drugih SCTP usluga korisnih za HTTP i standardizacija HTTP preko SCTP.
- ◆ Posledica su SCTP-specifične izmene u skladištima izvornih kodova za Apache i Firefox.

Studija: Web za regione u razvoju



- ◆ Kašnjenja od nekoliko 100ms do nekoliko sekundi.
- ◆ Potiče od uskog propusnog opsega i velikog propagacionog kašnjenja tipičnih bežičnih veza VSAT/3G/GPRS.
- ◆ Velika stopa grešaka u prenosu podataka.
- ◆ Rešenje: HTTP preko SCTP.

Postavka eksperimenata (1/2)



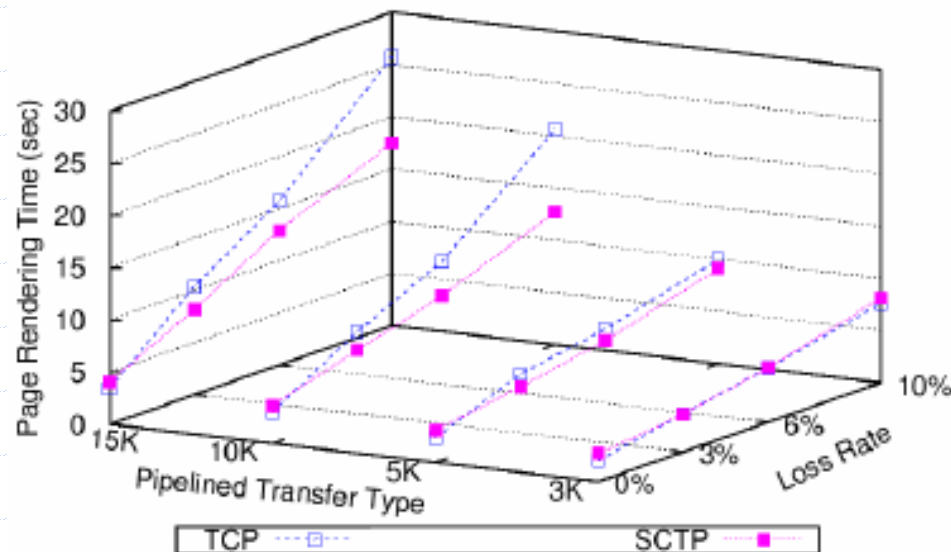
- ◆ Firefox klijent, Apache server i Dummynet čvor sa velikim redom čekanja između njih radi uobličavanja saobraćaja.
- ◆ Stopa grešaka od 0% do 10% (tipično za razmatrano okruženje).
- ◆ Brzina prenosa 1Mb/s.

Postavka eksperimenata (2/2)

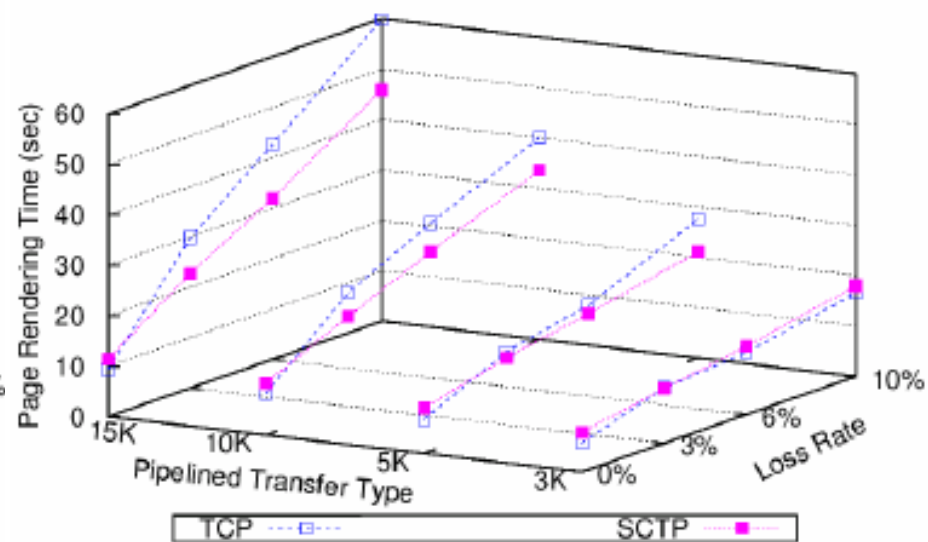
- ◆ Poredi se HTTP 1.1 perzistentan prenos podataka sa protočnom obradom preko TCP veze i SCTP veze sa više tokova.
- ◆ Protočnom obradom prenose se objekti iste veličine, i to veličine: 3KB, 5KB, 10KB i 15KB.
- ◆ Menjan se i broj objekata (N) prenošenih protočnom obradom. N se bira kao: 5, 10 ili 15.
- ◆ Mereno je:
 - Vreme potrebno za iscrtavanje cele stranice i
 - Vreme potrebno za iscrtavanje zadatog procenta stranice (pPage)

Vreme potrebno za crtanje cele stranice

1Mbps.350ms; N=10

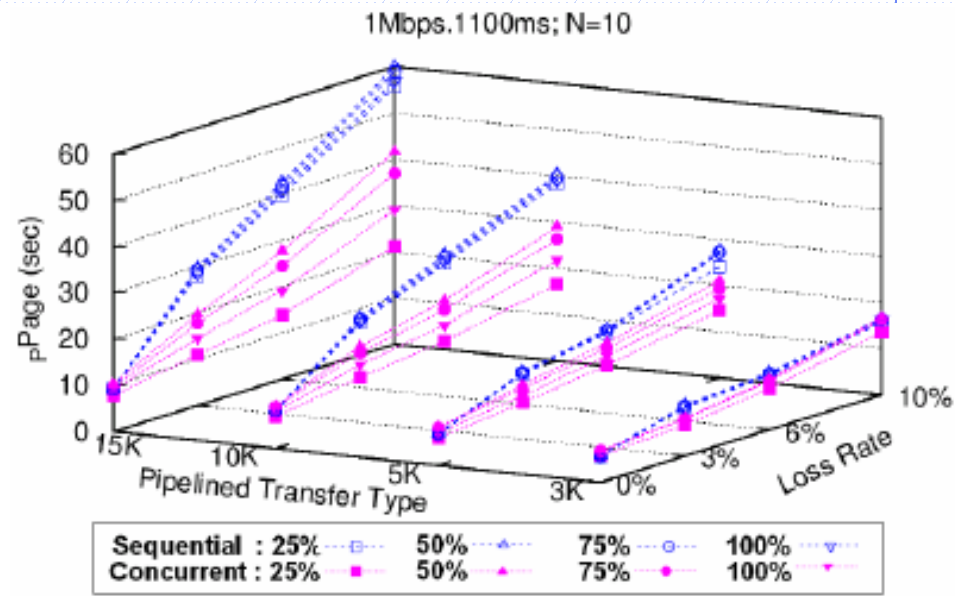
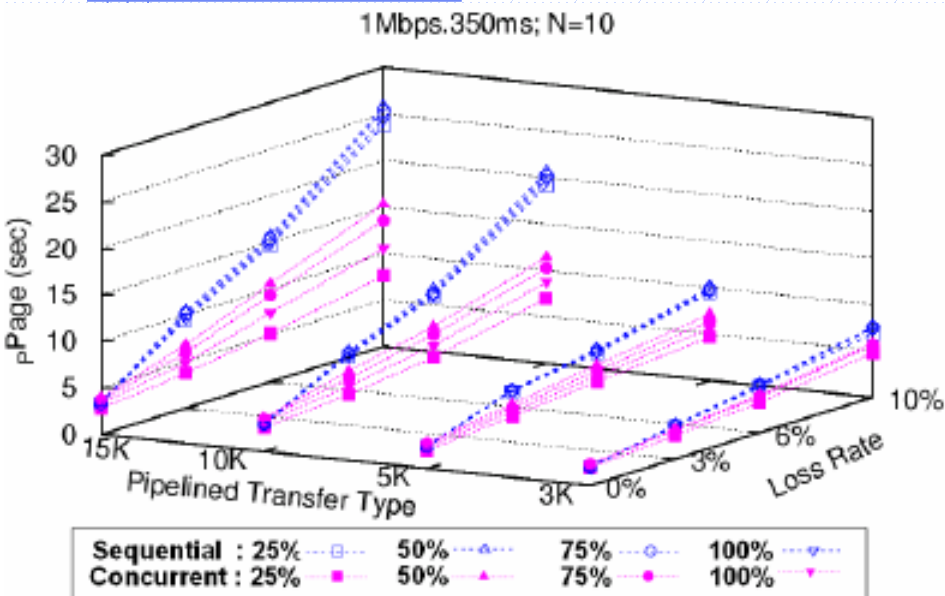


1Mbps.1100ms; N=10



- ◆ U oba slučaja N=10.
- ◆ Kašnjenje za sliku levo je 350ms, a za sliku desno 1100ms.
- ◆ TCP bele tačke, SCTP ljubičaste.

Vreme potrebno za crtanje zadatog procenta cele stranice (pPage)



- ◆ U oba slučaja N=10.
- ◆ Kašnjenje za sliku levo je 350ms, a za sliku desno 1100ms.
- ◆ TCP bele tačke, SCTP ljubičaste.

Proširenja API-ja utičnice (socket) (1/2)

- ◆ Sctp nudi većinu TCP usluga, ali i uključuje semantiku bližu UDP.
- ◆ Zbog toga postoje 2 modela Sctp utičnica.
 - Model jedan-na-jedan. Utičnica komunicira sa jednom Sctp krajnjom tačkom. Model sličan TCP za koji se TCP aplikacije lako mogu preraditi.
 - Model jedan-na-više. Utičnica komunicira sa više Sctp krajnjih tačaka, slično kao UDP utičnica.
- ◆ Sctp podržava sve postojeće API pozive, kao što su bind, listen, connect, read i write.
- ◆ Međutim postojeći API je neefikasan i ne može da iskoristi prednosti novina u Sctp.

Proširenja API-ja utičnice (socket) (2/2)

- ◆ Primeri novih API funkcije specifičnih za SCTP su:
 - `sctp_bindx` omogućava aplikaciji da poveže SCTP utičnicu sa više IP adresa.
 - `sctp_connectx` omogućava aplikaciji povezivanje sa drugom stranom (peer) korišćenjem više IP adresa.
 - `sctp_sendmsg` ima specijale argumente za slanje podataka određenim tokom, zadavanje zastavica za isporuku van redosleda, zadavanje vremena života za poruke sa delimičnom pouzdanošću, itd.
- ◆ SCTP događaji omogućavaju aplikaciji da reaguje na promene stanja mreže, npr. promena adrese druge strane (peer), greške u radu, itd.

Standardizacija

- ◆ IETF radna grupa za oblast transporta (TSVWG) je zadužena za održavanje SCTP.
- ◆ U toku je standardizacija modularnih proširenja, proširenja API-ja i poboljšanja u vezi zaštite.
- ◆ Aktivnosti unutar IETF Behave radne grupe uključuju:
 - Specifikaciju NAT (Network Address Translation) ponašanja za SCTP saobraćaj.
 - SCTP preko UDP inkapsulaciju da bi se omogućio SCTP prelaz preko prolaznih čvorova koji ne podržavaju SCTP.
- ◆ SCTP realizacije u jezgrima OS postoje za AIX, HP-UX, Linux, Solaris, itd. i kao drajveri za različite verzije Windows OS.