



Faglig kontakt under eksamen:
Bjarne Helvik (73 592667)

EKSAMEN I EMNE TTM4110 PÅLITELIGHET OG YTELSE MED SIMULERING

Dato: Tirsdag 30. november 2004

Tid: 1500 – 1900

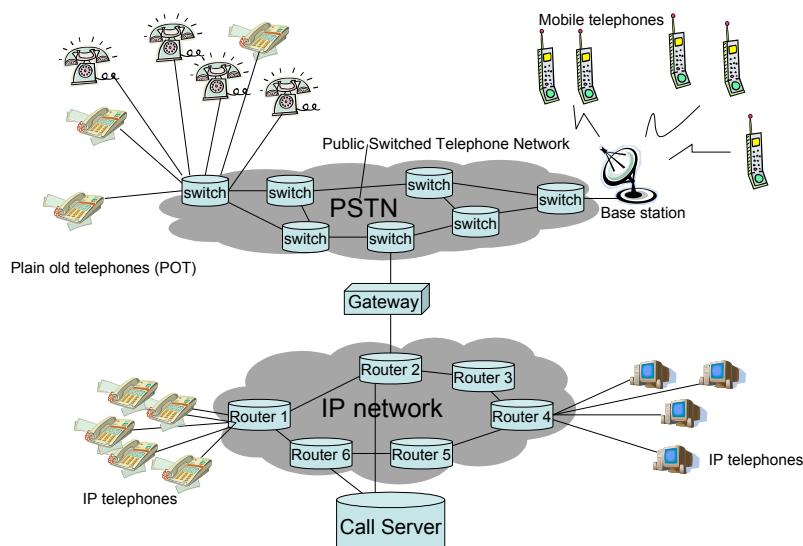
Hjelpemidler: C¹

Sensur: uke 1, 2005²

English text on page 6. If there are any differences between the Norwegian and English versions in semantics, the Norwegian version is superior.

¹Graham Birtwistle: DEMOS - A system for Discrete Event Modelling on Simula. (Personlige annotering i DEMOS bok er tillatt). Formelsamling i fag TTM4110 Pålitelighet og ytelse med simulering. NB! Formelsamlingen er vedlagt

²Merk! Studentene må primært gjøre seg kjent med sensur ved å oppsøke sensuroppslagene. Evt. telefoner om sensur må rettes til sensurtelefonene. Eksamenskontoret vil ikke kunne svare på slike telefoner.



Figur 1: IP telefoni arkitektur

Et firma kalt *TelefonIP* tilbyr en IP telenfontjeneste. Kundene har en IP-telefon hvorfra de kan ringe til, og motta samtaler fra andre IP-telefoner, vanlige fast-telefoner og mobiltelefoner. Den som ringer kalles A-side (originerende) mens den man ringer til er B-side (terminerende). I figur 1 er det vist en skisse av de ulike elementene som inngår i en slik tjeneste.

En samtale består av 3 faser (beskrivelsen antar at A- og/eller B-side er kunde av *TelefonIP*)³:

1. *Oppkopling* - A-siden sender B-nummeret til en samtale tjener (Call Server) i IP-nettet som returnerer IP-adressen til B-side. Hvis B-side er i PSTN returneres adressen til samtaleporten og en talekanal settes opp fra samtaleporten til B-side.
2. *Samtale* - A-side og B-side kan snakke sammen ved at informasjonen (talen) går pakkesvitsjet gjennom IP-nettet (Voice over IP) hvis begge er i IP-nettet. Hvis A- eller B-side er i PSTN så går talen linjesvitsjet i PSTN over en talekanal, og oversettes i samtaleporten til pakkesvitsjet tale. Oversettelsen krever en oversetterressurs som allokeres i samtaleporten. Tilsvarende blir pakkesvitsjet tale oversatt til linjesvitsjet tale ved bruk av en oversetter-ressurs i samtaleporten.

³Beskrivelsene er sterkt forenklet. Begrepene som er brukt i beskrivelsen er nøytrale i forhold til begreper som brukes i beskrivelse av SIP og H.323.

3. *Nedkopling* - Samtaletjeneren gis beskjed om at samtalen er over og varigheten registreres. Eventuelle ressurser bruk til talekanalen i samtaleport og PSTN frigis. Dette gjelder kun når A- eller B-side er i PSTN. Nedkoplingen kan initieres enten av A- eller B-side.

Samtaleporten (gateway) oversetter pakkesvitsjet tale til linjesvitsjet tale. Hver talekanal gjennom samtaleporten allokere en *oversetter-ressurs* i denne. Oppsetningsforsøk av talekanal når alle oversetterressursene er i bruk vil bli avvist og samtalen blir ikke oppkoplet.

Samtaletjeneren (Call Server) oversetter et telefonnummer til en IP adresse. I tillegg vil samtale-tjeneren registrere varigheten av hver samtale. Dette brukes som underlag i telefonregningen.

Det er åpenbart at kvaliteten til telefonitjenesten er avhengig av at ruterne, samtale-tjeneren (Call Server) og samtaleporten (gateway) har god ytelse og høy pålitelighet. I det etterfølgende skal vi for enkelhetsskyld anta at alle linjer, alle telefonsvitsjer og basestasjoner i PSTN, og alle terminaler, er feilfrie og ikke er ytelsesflaskehals.

- a) Definer tjenestekvalitetskrav til alle 3 faser i IP-telefonitjenesten både med hensyn på ytelses- og pålitelighetsegenskaper.

Anta at alle IP telefoner er koplet enten til ruter 1 eller ruter 4 som vist i figur 1. Tilgjengeligheten til en ruter er A_r (samme for alle rutere), til samtaleporten, A_{gw} , og samtaletjeneren, A_{st} .

- b) Gjør nødvendige antakelser og lag et pålitelighetsblokkskjema for IP-telefonitjenesten hvor A-side er koplet til ruter 1 og B-side er koplet til ruter 4. Betrakt *både* oppkopplings- og samtalefasen samtidig i samme modell. Sett opp et uttrykk for stasjonært tilgjengeligheten til IP-telefonitjenesten begrenset til IP-nettet. Definér hva et minimum stisett er og vis dette med to eksempler fra ditt pålitelighetsblokkskjema. Hva kan du kvalitativt si om påliteligheten til systemet utfra disse stisettene?

Påliteligheten til samtaletjeneren er kritisk og skal studeres nærmere. Anta at ved hver feil så avbrytes alle samtaler og alle samtaledata siden forrige fakturering slettes. Begge deler er svært uheldig og det er derfor viktig at tjeneren har høy funksjonsikkerhet. Det er kritisk for operatøren at alle samtaledataene siden forrige fakturering slettes fordi dette betyr at all inntekt fra avvirket trafikk er tapt for operatøren. Operatøren vil derfor vurdere hyppigheten av fakturering. Dette er en avveining mellom hva fakturering koster operatøren, C_b , og hvor mye inntekter som tapes hvis en feil skjer før fakturering er gjennomført. Anta at forventet samtaletid er T_{ab} , samtaleintensitet (uendelig brukergruppe), λ_{ab} , og pris per samtale er c_{ab} . Hvis kostnadene ved fakturering er større enn forventet tapt inntekt så fakturerer man ikke.

- c) Beregn funksjonssannsynligheten til samtaletjeneren når vi antar at feilingsraten er λ . Sett opp uttrykket for MTTF og MTFF for samtaletjeneren. Bruk funksjonssannsynligheten til å sette opp et uttrykk som bestemmer minimumsintervallet mellom fakturering. Uttrykket skal ikke løses.

Tabell 1: Trafikkinteressematrise [angitt i Erlang].

A-side \ B-side	til IP telefon	til mobiltelefon	til fasttelefon (POT)
fra IP-telefon	140	27	38
fra mobiltelefon	19	150	80
fra fasttelefon (POT)	40	90	200

Ytelsen til samtaleporten er identifisert som en annen flaskehals med hensyn på tjenestekvaliteten til deler av telefonitjenesten. Anta at trafikkinteressematrisen mellom de ulike telefoni-typene er som i tabell 1. Tabellen inneholder tilbudt trafikk fra begge selskapene til *TelefonIAS*, nemlig *TelefonIP* og *TelefonI2* hvor sistnevnte leverer telefoni til kunder i PSTN.

- d) Hva er tilbudt trafikk mellom kunder av TelefonIP og TelefonI2, dvs. i begge retninger over samtaleporten? Hva blir tidssperr for anrop over samtaleporten med 128 oversetter-ressurser? Under hvilke antagelser kan du bruke Erlangs tapsformel som angitt i tabell 5? Hvor mange oversetter-ressurser må samtaleporten ha hvis tilbudt trafikk er 140 erlang for at anropssperr skal være mindre enn 10%? Hvis tilbudt trafikk øker til 162 erlang uten at antall ressurser økes, hva blir avvirket trafikk over samtaleporten da?

For å dimensjonere samtaleporten og vurdere feiltolerant design av samtaltjeneren skal sannsynligheten for at en samtale blir avvist (sperret) evalueres ved hjelp av simulering.

- e) Beskriv de situasjoner hvor oppsett av en samtale mislykkes og kommenter hvilke situasjoner du tar med i modellen din. Antagelsene gitt i den generelle beskrivelsen er fortsatt gyldige. Foreslå entiteter og ressurser for en simuleringsmodell som kan evaluere sannsynligheten for anropssperr for IP-telefonitjenesten. Skissér ved hjelp av aktivitetsdiagrammer en simuleringsmodell for dette. Simuleringsmodellen skal avgrenses til IP-telefoni mellom A- og B-side hvor begge er i IP-nettet (dvs. kunder av *TelefonIP*).

Tabell 2: Samtalevarighet [minutter].

Observasjon	k	Samtale IP-IP	Samtale IP-mobil	Samtale IP-fast
$\sum_{i=1}^{n_k} X_{ik}$	1	879	134	245
$\sum_{i=1}^{n_k} X_{ik}^2$	2	2989	270	673
n_k	3	300	90	105

Simuleringen har registrert samtalevarighet for alle samtaler som settes opp via samtaletjeneren. Anta at X_{ik} er varighet av samtale i av type $k = 1, 2, 3$. I tabell 2 er tallene gitt for summert varighet, $\sum_{i=1}^{n_k} X_{ik}$, summen av kvadratene, $\sum_{i=1}^{n_k} X_{ik}^2$, og antall observasjoner for hver type, n_k .

f) Estimér forventet samtaletid og variansen for alle samtaler.

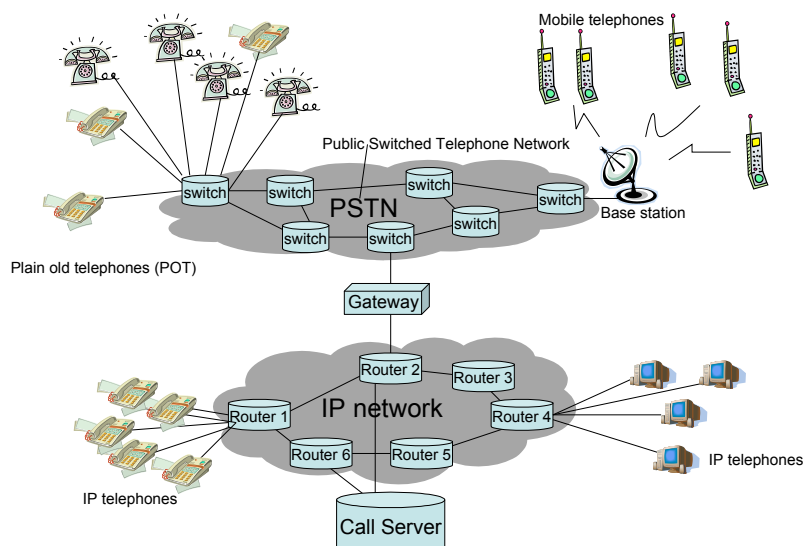
Basert på simuleringene har man funnet ut at det ikke er tilstrekkelig med én samtaletjener. For å forbedre påliteligheten så dubleres samtaletjeneren (dvs. man har to *aktive* tjenere) og at vi i tillegg har en varm reserve (en *passive*, men oppdatert tjener). Tjenesten virker så lenge minst én aktiv tjener er arbeidende. Anta at alle 3 tjenerne er identiske og at de har samme feilrate, λ , når de er aktive, og halve raten når de er passive, $\lambda/2$. Betrakter kun transiente programvarefeil. Feilet aktiv tjener erstattes med reserve hvis reserven er arbeidende. Forventet tid for å erstatte feilet aktiv med resevere er $1/\mu_s$. Hver maskin repareres med hver sin programvarestyrte restartprosess og feil oppdages med egen diagnostiseringsprosess som også kjører på hver maskin. Forventet er tid for en restart er $1/\mu_r$. Alle tider er negativt eksponenital fordelte.

g) Lag tilstandsmodell. Vis hvordan MTTF for samtaletjeneren kan beregnes.

Utilgjengeligheten for samtaletjeneren er

$$U = \frac{\lambda^2(100\lambda\mu_r^2 + 48\mu_r^3 + 4\lambda\mu_r(17\lambda + 19\mu_s) + 5\lambda(3\lambda^2 + 8\lambda\mu_s + 4\mu_s^2))}{48\mu_r^5 + 4\mu_r^4(49\lambda + 18\mu_s) + 5\lambda^3(3\lambda^2 + 8\lambda\mu_s + 4\mu_s^2) + \mu_r^3(316\lambda^2 + 238\lambda\mu_s + 24\mu_s^2) + 2\lambda^2\mu_r(49\lambda^2 + 93\lambda\mu_s + 30\mu_s^2) + \lambda\mu_r^2(251\lambda^2 + 313\lambda\mu_s + 60\mu_s^2)}$$

h) Anta at $(\mu_s, \mu_r) \gg \lambda$. Finn et forenklet uttrykket for U og gi en kvalitativ vurdering av dette relatert til systemet og modellen.



Figur 2: IP telephony architecture

A firm called *TelefonIP* offers an IP telephony service. Their customers have an IP phone from which they can make a call, and receive calls from other IP phones, as well as ordinary fixed and mobile phones. The caller is denoted A-side, and receiver of a call is denoted B-side. In Figure 2 the elements in this service are illustrated.

A call consists of 3 phases (the description assumes that A- and/or B-side are/is a *TelefonIP* customer)⁴:

1. *Setup* - the A-side sends the B-number to a Call Server in the IP network. This returns the B-side address. If the B-side is in the PSTN the returned number is to the gateway and a speech channel is setup from the gateway to the B-side in PSTN.
2. *Call* - A- and B-side can speak by exchanging information (speech) packet-switched if both sides are in the IP network. If either A- or B-side is in the PSTN then the speech is circuit switched in the PSTN, translated in the gateway, to packet switched information. The translation requires and allocates a translation resource in the gateway.

⁴The descriptions are simplified. The terms used in the descriptions are neutral relative to terms from SIP and H.323.

3. *Disconnection* - either A- or B-side notifies the Call Server that the call is completed and the call duration is recorded. If a translation resource is allocated in the gateway this is freed.

The gateway translates between packet switched and circuit switched speech. Every speech channel connected through the gateway allocates a translation resource. Call setup attempts when all resources are allocated will be rejected.

The Call Server translates a phone number into an IP address. In addition the Call Server will record the duration of every call. This information is the basis for the phone bill.

Obviously, the quality of the telephony service is dependent on high performance and dependability of routers, Call Server and gateway. For simplicity, in the following we will assume that all trunk lines, all switches and base stations in the PSTN, and all terminal never fail and with no performance constraints.

- a) Specify service quality level requirements for all 3 phases of the IP telephony service with respect to both performance and dependability.

Assume that all IP phones are connected to either router 1 or 4 as shown in Figure 2. The availability of a router is A_r (the same for all routers), of the gateway, A_{gw} , and of the Call Server A_{st} .

- b) Make the necessary assumptions and draw a reliability block diagram (no: pålitelighets-blokkskjema) for the IP telephony service where A-side is connected to router 1 and B-side is connected to router 4. Consider both setup and call phases simultaneously in the model. Give an expression of the steady state availability of the IP telephony service within the IP network. Define a minimum path set (no: minimum stisett) and demonstrate this by two examples from your reliability block diagram. What can you qualitatively deduce from these path sets with respect to the system dependability?

The dependability of the Call Server is critical and will be studied more closely. Assume that on failure every ongoing call will be interrupted, and all recorded billing data since last invoice will be lost. Both consequences are serious. Hence, it is important that the reliability (no: funksjonsikkerhet) of the server is high. It is very critical for the operator if all billing data are lost because this means that all income since last invoice will be lost. The operator must consider how often the invoice should be made. This is a trade off between potential loss of income and the expenses, C_b , related to generating the invoice. Assume that expected call duration is, T_{ab} , call arrival intensity (infinite number of subscribers), λ_{ab} , and the cost per call is c_{ab} . If the cost of billing (generating the invoice) exceeds the expected loss of income, the operator will not invoice.

- c) Determine the reliability (no: funksjonssannsynligheten) of the Call Server when expected failure rate is constant λ . Express the MTTF and MTBF for the Call Server. Apply the reliability function (no: funksjonssannsynligheten) to determine the minimum interval between billing. The expression should not be solved.

Tabell 3: Traffic intensity matrix [given in Erlang].

A-side \ B-side	to IP phone	to mobile phone	to fixed phone (POT)
from IP phone	140	27	38
from mobile phone	19	150	80
from fixed phone (POT)	40	90	200

The performance of the gateway is identified as a bottleneck of the quality of the IP telephony service. Assume that the traffic intensity matrix (no: trafikkinteressematrisen) between the various customers are as given in Table 3. The table contains offered traffic of both companies of *TelefonIAS*, namely *TelefonIP* and *TelefonI2* where the latter offered telephony service to customers in the PSTN.

- d) What is offered traffic between customers of *TelefonIP* and *TelefonI2*, i.e. in both directions over the gateway? What is time congestion (no: tidssperr) over the gateway with 128 translation resources? Under what assumptions can Erlang's formula in Table 5 be applied? How many translation resources is required given offered traffic of 140 erlang, and call congestion (no: anropssperr) of less than 10%? If the offered traffic increases to 162 erlang without increasing the number of resources, what is the carried traffic (no: avviklet trafikk) over the gateway?

To dimension the gateway and evaluate a fault tolerant design of the Call Server, the probability of call congestion should be estimated by simulation.

- e) Describe the different causes of call setup failures and comment which cases you will include in your model. The assumptions given in the introduction are still valid. Propose entities and resources for a simulation model where call congestion for the IP telephony service can be estimated. Draw your simulation model by use of activity diagrams. The simulation model should be limited to the IP telephony service between A- and B-side where both are in the IP network (i.e. customers of *TelefonIP*)

Tabell 4: Call durations [minutes].

Observations	k	Call IP-IP	Call IP-mobile	Call IP-fix
$\sum_{i=1}^{n_k} X_{ik}$	1	879	134	245
$\sum_{i=1}^{n_k} X_{ik}^2$	2	2989	270	673
n_k	3	300	90	105

From the simulations the call durations for all calls via the Call Server are observed. Assume that X_{ik} is the call duration of call i of type $k = 1, 2, 3$. Table 4 contains the sum of call duration, $\sum_{i=1}^{n_k} X_{ik}$, the sum of squared call duration, $\sum_{i=1}^{n_k} X_{ik}^2$, and the number of calls of each type, n_k .

f) Estimate the expected call duration and the variance of all calls.

Based on the simulation it is discovered that one Call Server is not sufficient. To improve the dependability two *active* Call Servers are introduced. In addition we have one hot standby (*passive* but updated server). The telephony service is operational while at least one active server is operational. Assume 3 identical servers with the same failure rate, λ , when they are active, and half the rate when they are passive, $\lambda/2$. Only transient software failures are considered. A failed active server is substituted with a standby if the standby is operational. Expected time to make the standby active is $1/\mu_s$. Each machine is repaired by individual software-controlled restart process and failures are detected by a diagnostic process that runs on each machine. Expected restart time is $1/\mu_r$. All times are negative exponentially distributed.

g) Make a state diagram model (no: tilstandsmodell). Explain how the MTTF for the Call Server can be obtained.

The unavailability of the Call Server is

$$U = \frac{\lambda^2(100\lambda\mu_r^2 + 48\mu_r^3 + 4\lambda\mu_r(17\lambda + 19\mu_s) + 5\lambda(3\lambda^2 + 8\lambda\mu_s + 4\mu_s^2))}{48\mu_r^5 + 4\mu_r^4(49\lambda + 18\mu_s) + 5\lambda^3(3\lambda^2 + 8\lambda\mu_s + 4\mu_s^2) + \mu_r^3(316\lambda^2 + 238\lambda\mu_s + 24\mu_s^2) + 2\lambda^2\mu_r(49\lambda^2 + 93\lambda\mu_s + 30\mu_s^2) + \lambda\mu_r^2(251\lambda^2 + 313\lambda\mu_s + 60\mu_s^2)}$$

h) Assume that $(\mu_s, \mu_r) \gg \lambda$. Establish a simplified expression of U and give a qualitative validation of this related to the system and the model.

Tabell 5: Utdrag fra “Tabell for Erlangs tapsformel”

Tallene oppgir tilbudt trafikk A gitt et antall organ n og en tapssannsynlighet, E .

n = 101 - 151

Offered traffic flow A in erlang

n	Loss probability (E)										n
	0.007	0.008	0.009	0.01	0.02	0.03	0.05	0.1	0.2	0.4	
101	83.296	83.916	84.482	85.005	88.941	91.784	96.265	105.21	121.89	165.92	101
102	84.225	84.849	85.419	85.946	89.910	92.774	97.289	106.30	123.13	167.58	102
103	85.155	85.783	86.357	86.888	90.880	93.765	98.314	107.40	124.38	169.25	103
104	86.086	86.718	87.296	87.830	91.850	94.756	99.339	108.50	125.63	170.91	104
105	87.017	87.653	88.235	88.773	92.821	95.747	100.36	109.60	126.87	172.58	105
106	87.948	88.589	89.175	89.716	93.791	96.738	101.39	110.70	128.12	174.25	106
107	88.880	89.525	90.115	90.660	94.763	97.730	102.42	111.79	129.36	175.91	107
108	89.812	90.462	91.055	91.604	95.734	98.722	103.44	112.89	130.61	177.58	108
109	90.745	91.399	91.996	92.548	96.706	99.715	104.47	113.99	131.86	179.24	109
110	91.678	92.336	92.937	93.493	97.678	100.71	105.49	115.09	133.10	180.91	110
111	92.612	93.274	93.879	94.438	98.651	101.70	106.52	116.19	134.35	182.58	111
112	93.546	94.212	94.821	95.384	99.624	102.69	107.55	117.29	135.59	184.24	112
113	94.481	95.151	95.764	96.330	100.60	103.69	108.57	118.39	136.84	185.91	113
114	95.416	96.090	96.707	97.277	101.57	104.68	109.60	119.49	138.09	187.57	114
115	96.352	97.030	97.650	98.223	102.54	105.68	110.63	120.58	139.33	189.24	115
116	97.287	97.970	98.594	99.171	103.52	106.67	111.66	121.68	140.58	190.91	116
117	98.224	98.910	99.538	100.12	104.49	107.66	112.69	122.78	141.83	192.57	117
118	99.160	99.851	100.48	101.07	105.47	108.66	113.71	123.88	143.07	194.24	118
119	100.10	100.79	101.43	102.01	106.44	109.66	114.74	124.98	144.32	195.91	119
120	101.04	101.73	102.37	102.96	107.42	110.65	115.77	126.08	145.57	197.57	120
121	101.97	102.68	103.32	103.91	108.39	111.65	116.80	127.18	146.81	199.24	121
122	102.91	103.62	104.26	104.86	109.37	112.64	117.83	128.28	148.06	200.90	122
123	103.85	104.56	105.21	105.81	110.35	113.64	118.86	129.38	149.31	202.57	123
124	104.79	105.50	106.16	106.76	111.32	114.64	119.89	130.48	150.55	204.24	124
125	105.73	106.45	107.10	107.71	112.30	115.63	120.92	131.58	151.80	205.90	125
126	106.67	107.39	108.05	108.66	113.28	116.63	121.95	132.68	153.05	207.57	126
127	107.61	108.34	109.00	109.61	114.25	117.63	122.98	133.78	154.29	209.23	127
128	108.55	109.28	109.95	110.57	115.23	118.62	124.01	134.88	155.54	210.90	128
129	109.49	110.22	110.90	111.52	116.21	119.62	125.04	135.99	156.79	212.57	129
130	110.43	111.17	111.85	112.47	117.19	120.62	126.07	137.09	158.03	214.23	130
131	111.37	112.12	112.79	113.42	118.17	121.62	127.10	138.19	159.28	215.90	131
132	112.31	113.06	113.74	114.38	119.15	122.62	128.13	139.29	160.53	217.57	132
133	113.26	114.01	114.69	115.33	120.12	123.61	129.16	140.39	161.77	219.23	133
134	114.20	114.95	115.64	116.28	121.10	124.61	130.19	141.49	163.02	220.90	134
135	115.14	115.90	116.59	117.24	122.08	125.61	131.22	142.59	164.27	222.56	135
136	116.09	116.85	117.54	118.19	123.06	126.61	132.25	143.69	165.52	224.23	136
137	117.03	117.80	118.50	119.14	124.04	127.61	133.28	144.80	166.76	225.90	137
138	117.97	118.74	119.45	120.10	125.02	128.61	134.32	145.90	168.01	227.56	138
139	118.92	119.69	120.40	121.05	126.00	129.61	135.35	147.00	169.26	229.23	139
140	119.86	120.64	121.35	122.01	126.98	130.61	136.38	148.10	170.50	230.90	140
141	120.81	121.59	122.30	122.96	127.97	131.61	137.41	149.20	171.75	232.56	141
142	121.75	122.54	123.26	123.92	128.95	132.61	138.44	150.30	173.00	234.23	142
143	122.70	123.49	124.21	124.88	129.93	133.61	139.48	151.41	174.25	235.89	143
144	123.64	124.44	125.16	125.83	130.91	134.61	140.51	152.51	175.49	237.56	144
145	124.59	125.39	126.11	126.79	131.89	135.61	141.54	153.61	176.74	239.23	145
146	125.54	126.34	127.07	127.75	132.87	136.61	142.57	154.71	177.99	240.89	146
147	126.48	127.29	128.02	128.70	133.86	137.61	143.61	155.82	179.24	242.56	147
148	127.43	128.24	128.98	129.66	134.84	138.61	144.64	156.92	180.48	244.23	148
149	128.38	129.19	129.93	130.62	135.82	139.62	145.67	158.02	181.73	245.89	149
150	129.32	130.14	130.88	131.58	136.80	140.62	146.71	159.12	182.98	247.56	150
151	130.27	131.09	131.84	132.53	137.79	141.62	147.74	160.23	184.23	249.22	151
n	0.007	0.008	0.009	0.01	0.02	0.03	0.05	0.1	0.2	0.4	n