



Budapesti Műszaki és Gazdaságtudományi Egyetem

Villamosmérnöki és Informatikai Kar

Távközlési és Médiainformaticai Tanszék

Szabály alapú útvonalépítési stratégiák nagy hálózatokban

DIPLOMATERV

Készítette
Döbrei Gábor

Konzulens
dr. Heszberger Zalán

2014. április 21.

HALLGATÓI NYILATKOZAT

Alulírott *Döbrei Gábor*, szigorló hallgató kijelentem, hogy ezt a diplomatervet meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy autentikált felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Budapest, 2014. november 30.

Döbrei Gábor
hallgató

Köszönetnyilvánítás

Nagyon sokat köszönhetek dr. Heszberger Zalánnak, aki alapdiplomás szakdolgozati, majd mesterdiplomás konzulensként nem csak elindított a kutatási irányba, de menet közben sok konzultációval, és baráti vezetéssel segítette fejlődésemet. A közös munka inspirációt is jelentett és megismertetett sok olyan eredménnyel, cikkel, amelyre egymagam nem találtam volna rá.

Sok közvetett segítséget és érdeklődést kaptam barátaimtól, évfolyamtársaimtól. Végül családom, elsősorban feleségem végtelen türelmére szükség volt a dolgozat létrejöttéhez.

Tartalomjegyzék

Köszönnetnyilvánítás	2
Kivonat	4
Bevezető	5
1. Általánosított hálózatok és útvonalválasztási szabályok	8
1.1. Történeti áttekintés	8
1.2. Útvonalválasztás, routing algebrák	10
1.3. Routing algebrák tulajdonságai	13
1.3.1. Monotonitás, izotonitás és egyéb tulajdonságok	13
1.4. Műveletek routing algebrák között	14
1.4.1. A műveletek hatása a tulajdonságokra	15
1.5. Példák algebráakra és kombinálásukra	15
1.6. Összefoglaló	16
2. A hálózatkutatás legfontosabb modelljei	17
2.1. Vírusterjedés komplex hálózatokban	18
2.1.1. A vírusterjedés matematikai modellje	18
2.1.2. Vírusterjedés, mint útvonalválasztási probléma	19
2.2. Trendterjedés közösségi hálózatokban	21
2.2.1. Trendterjedés, mint útvonalválasztási probléma	21
2.3. Útvonalválasztás az Interneten	23
2.4. Egyéb algebrák	26
2.5. Összefoglaló	27
3. A modellezési keretrendszer	28
3.1. Specifikáció	28
3.1.1. A bemenet	28
3.1.2. A szimuláció	28
3.1.3. A kimenet	28
3.2. A szimulátor	28
3.2.1. A matematikai struktúrák implementálása	28
3.2.2. A felhasznált algoritmusok elemzése	28

3.3.	A szimulációs eredmények kiértékelése	28
3.3.1.	Az eredmények validálása	28
3.3.2.	Gráfok összehasonlítása	28
3.3.3.	Az adatok megjelenítése	28
3.4.	Valós hálózatok vizsgálata a keretrendszer segítségével	28
3.5.	Összefoglaló	29
4.	Repülési hálózat vizsgálata	30
4.1.	Az adatok	30
4.1.1.	Repterek	30
4.1.2.	Repülési útvonalak	32
4.1.3.	Repülési társaságok	33
4.2.	A szimuláció	33
4.2.1.	Az adatok előfeldolgozása	33
4.2.2.	A szimuláció megtervezése	33
4.2.3.	A vizsgálandó metrikák, az élek súlyozása	34
4.3.	Az eredmények	34
4.4.	Összefoglaló	34
5.	Összefoglalás	35
Ábrák jegyzéke		36
Irodalomjegyzék		37
Függelék		39
F.1.	Számításelméleti alapok	39
F.2.	Az útvonalválasztás teljes modellje	40
F.3.	Tételek a routing algebrák témakörében	40
F.4.	A szimulátor technikai részletei	41
F.4.1.	A választott platform	41
F.4.2.	Fejlesztői megjegyzések	41

Kivonat

Az Internet rohamos fejlődésével az utóbbi időben mind nagyobb hangsúlyt kapnak a hagyományostól eltérő hálózatmenedzsment funkciókat megvalósító algoritmus kutatások. Ezek egyik legaktívabban művelt ága az útvonalválasztás kérdéseivel foglalkozik, és lényegében az erőforrások optimalizálást célzó klasszikus (főként legrövidebb utak megtalálására koncentrált) algoritmusok alternatíváinak kutatását tűzi ki célul. Az algoritmusok általános szabályok (policy) mentén alakítják ki a lehetséges kommunikációs útvonalakat.

A Diplomaterv célja olyan szabályalapú útvonalépítési stratégiák kutatása (szintetizálása/elemzése), mely mindenmellett, hogy alkalmas lehet valós kommunikációs hálózati alkalmazásra is, felhasználható más természetes vagy mesterséges/technológiai valós hálózatok működésének felderítésére.

Az első feladatom a Diplomaterv készítése során áttekinteni a hálózatok és az útvonalválasztás matematikai modellezési lehetőségeit. Másodszor, eddig még nem vizsgált szabályokat definiálok és szimulációs módszerekkel vizsgálok meg, illetve a valós hálózatok útvonalválasztásával hasonlítom össze. Végül javaslatot teszek, hogy hogyan lehet felhasználni a számítógépes hálózatok útvonalválasztásában az egyéb tudományterületekről származó problémák vizsgálat során szerzett új ismereteket.

Bevezető

Az Internet rohamos fejlődésével egyre nagyobb igény mutatkozik a komplex számítógépes hálózatok megismerésére. Általános esetben a (kis) hálózatok struktúrája és működési mechanizmusai jól meghatározottak, hiszen a saját tervezésünk eredményeként jöttek létre és a vezérlés is a mi kezünkben van. Pontosan tudjuk, hogy egy hálózati csomópont melyik másik csomóponttal van kapcsolatban, ismerjük az összeköttetéseket és az útvonalválasztást meghatározó szabályokat is. Ha bármilyen módosítást szeretnénk eszközölni, vagy egy hibát szeretnénk kijavítani, azonnal (nagyon gyorsan) tudjuk, hogy hol kell beavatkozni. Azonban egy olyan hálózat vizsgálata során, amelyet nem mi tervezünk, nagyon gyorsan szembesülünk olyan kérdésekkel, amiket csak nehezen és sok munka árán tudunk megválaszolni - mérésekkel, teszteléssel - és csak abban az esetben, ha a hálózat mérete még nem jelent problémát.

Az útvonalakat meghatározó algoritmusok általános szabályok (policy) mentén alakítják ki a lehetséges (kommunikációs) útvonalakat és bár a kapcsolódó irányadó paraméterek igen változatosak lehetnek, a problémakör egy meghatározó paramétere szerinti legrövidebb útvonalat szoktuk a legjobbnak tekinteni. Mégis, ha megvizsgáljuk az Internet magas szintű topológiáját¹ és a benne kialakult utakat, akkor azt vesszük észre, hogy ezek az utak nem az optimális megoldások, legalábbis nem a legrövidebbek. Mivel feltehetjük, hogy az anyagi haszon maximalizálása céljából az Autonóm Rendszerek² üzemeltetői racionális döntések révén építették ki pontosan ezeket az utakat, jogos kérdés, hogy pontosan milyen stratégia alapján tették ezt. Mi vezethet egy látszólag ésszerűtlen döntéshez? Általánosságban igaz az egyéb hálózatokra is, hogy az adott szituáció optimális útjai nem tűnnek racionálisnak.

Ha körbenézünk a világban, az informatikától távoleső területeken is rengeteg példát találunk olyan hálózatokra, amiknek nem értjük még a működését, nem tudjuk pontosan leírni a belső folyamatait. Számos példát találunk a biológiából, a szociológiából vagy a pénzügyi világból, ugyanakkor minden ilyen probléma vizsgálatát vissza lehet vezetni egy olyan modell vizsgálatára, ami általában képes kezelni magát a hálózat fogalmát és az azon történő útvonalak kialakulását/kialakítását. A vírusok által terjesztett betegség terjedése, egy ruhaviselet divattá válása, és a számítógépes hálózatok kommunikációs útjainak kialakulása modellezhető a gráfelmélet eszközrendszerével. Mindhárom esetben jól meg-

¹Nevezik még az Internet tartomány-szintű-, vagy AS-szintű topológiának, gráfjának is.

²Autonóm Rendszer - Autonomous System (AS)- : önálló útválasztási tartomány, amelyen belül egyetlen, jól meghatározott útvonalválasztási szabály érvényesül.

határozható, elkülönülő csomópontok vannak, akik között terjed egy csomag, ami lehet információ vagy pl. a betegség maga. Egy modellen belül a csomópontok általában nem különböznek egymástól, és a köztük levő kapcsolatok különböző tulajdonságokkal rendelkeznek, amit az útvonalak kialakulása közben figyelembe is kell venni. Nem tételezünk fel különbséget két ember között, bárki meg tud betegedni. Magától értetődőnek látszik ugyanakkor, hogy levegőben terjedő vírusos fertőzés jóval lazább kapcsolaton keresztül is továbbterjed, míg egy vér útján terjedő betegség sokkal szorosabb kapcsolaton tud csak továbbterjedni: a kapcsolatokat súlyozni kell. Ha egy ember fogékonyabb egy betegségre, akkor a hozzá tartozó éleken nagyobb valószínűséggel terjed tovább majd a betegség. Ugyanilyen meggondolásból, két szomszédos router között a nagyobb sávszélességű úton továbbítjuk a csomagokat. Érdekes azonban, hogy egy divat elterjedését már nem tudjuk ilyen egyszerűen leírni, hiszen nem is igazán az a meghatározó, hogy mennyire befolyásos és ismert emberek reklámozzák, hanem az, hogy a társadalom felkészült-e már a befogadásra [1, 2].

A Diplomatervben a végső cél olyan policy primitíveket meghatározni, melyek összekapcsolásával jól lehet közelíteni a valós hálózatokat. Jelenleg sok megválaszolatlan kérdés van a BGP hálózat útvonalválasztásáról, melyek jó része onnan ered, hogy a hálózat csomópontjai, az AS-k nem fedik fel szabályrendszerüket. Ezen kérdésekre tudnánk válaszolni úgy, hogy nem a pontos szabályokat adjuk meg, hanem megmutatjuk, hogy adott policy primitívek felhasználásával majdnem ugyanolyan utakat kapunk, mint egy adott AS. Ez azért is lenne egy jó megoldás, mert lehet, hogy egy AS viszonylag bonyolult, sok erőforrást felemészítő szabályokat alkalmaz, pedig nagyon hasonló eredményeket érhetne el úgy is, ha az itt bemutatott p. primitíveket használná.

Az 1. fejezetben a számítógépes hálózatok policy-felderítésével foglalkozó szakirodalom áttekintés után egy általános hálózati- és routing modellt írok le, amely képes kezelni más tudományterületekről származó hasonló problémákat is. Az útvonalválasztás szabályrendszerét egy jól definiált matematikai struktúrával kell meghatározni, hogy definálni lehessen policy-k közti műveleteket, amelyekkel össze is tudunk kapcsolni policy-ket és tudjuk vizsgálni az kölcsönhatásukat.

A 2. fejezetben változatos problémákat és a hozzájuk tartozó routing policy-ket definiálunk és leírom a hozzájuk tartozó matematikai struktúrákat.

A 4. fejezetben szimulációs módszerekkel elemzem, hogy egyes teszt-hálózatokban milyen utakat határoznak meg az 1. fejezetben definiált policy-k, és javaslatot teszek, hogy egy valós hálózat útvonalválasztását melyik policy-k keverékével lehet legpontosabban modellezni. Ezután több valós hálózatot vizsgálok meg az addigra meghatározott saját policy-k segítségével, és megvizsgálom a keverék policy pontosságát, azaz azt, hogy mennyire térnének el a valós hálózatbeli utak a jelenlegitől akkor, ha az általam ismertetett policy-ket határoznánk meg azokat.

A feladatok elvégzéséhez a NetLogo³ nevű hálózati szimulált fogom használni, a valós hálózatok a BGP hálózat és egy repülési útvonalakat tartalmazó hálózat lesz.

A számításelmélet, az algoritmuselmélet és egyéb magasabb szintű témakörök ismerete elengedhetetlen ennek a diplomamunkának a megértése során, ugyanakkor minden fontosabb matematikai téTEL megtalálható a Függelékben.

³<http://ccl.northwestern.edu/netlogo/>

1. fejezet

Általánosított hálózatok és útvonalválasztási szabályok

Az Internet gerinchálózatát kialakító szabályrendszert nagy vonalaiban ismerjük. Az Internet AS-szintű topológiáját a BGP¹ határozza meg. Különböző eljárásokat ismerünk az Autonóm Rendszerek közti kapcsolatok feltárására a BGP-s routing táblák alapján, illetve az Internet router-szintű topológiájából, sőt az AS-ek routing-policy²-ját is tudjuk becsülni, ugyanakkor arra a kérdésre eddig még senki sem tudott választ adni, hogy miért éppen így alakultak ki a kommunikációs utak.

Ebben a fejezetben a szakirodalom rövid áttekintése után egy általános hálózati- és routing modellt mutatok be. Az útvonalválasztás szabályrendszerét egy jól definiált matematikai struktúrával - a routing algebrákkal - írom le, és definíálom a policy-k közti műveleteket, amelyekkel össze is tudunk kapcsolni policy-kat. Emellett egy alfejezet foglalkozik a policy-k különböző tulajdonságaival, melyek figyelembevétele fontos szempont a szimulációk megtervezésénél.

1.1. Történeti áttekintés

Az 1990-es évek végére a kutatók felismerték a tényt, hogy a BGP szintű Internet hálózatról nem tudunk szinte semmit. Rendelkezésükre állt néhány szolgáltató BGP-s routing táblája, de ez nagyon kevésnek bizonyult, hiszen a legtöbb AS közötti kapcsolat titkos gazdasági döntés révén született, így pontos képük nem lehetett a hálózatról. Ekkor kezdték el vizsgálni a különböző lehetőségeket, hogy hogyan lehetne feltárni ezt a rejtett hálózatot. Nem tűnt reménytelennek a helyzet, hiszen hálózat széle - a végpontok - nem tartoznak a szolgáltatókhöz, így egy tetszőleges számítógépről indított csomag útját végigkövetve értékes információkat nyerhettek. Természetesen a *traceroute*³ futási eredményeit elemezve egy halom adatot kapunk, amit nehezen lehet csak feldolgozni, annál is inkább, mert

¹Border Gateway Protocol

²Útvonalválasztási szabályrendszer

³Egy számítógép-hálózati diagnosztikai eszköz, amivel az IP (Internet Protokoll) hálózaton haladó csomagok útját lehet követni.

az egyes mérések ismétlése során más és más útvonalon haladt át a követendő csomag. Készültek szimulátorok, amik ügyesen kezelték a nehézségeket, és különböző heurisztikákat használtak, hogy csökkentsék a szükséges mérések számát [3], ugyanakkor még mindig nagy szakadék volt az eredmények és az elvárások között. Nem volt egyértelmű, hogy egy csomag miért épp az adott AS-eken keresztül ért célba.

Az ezredforduló után már sok új kutatás foglalkozott az Internet AS-szintű topológiájának feltérképezésével úgy, hogy az ehhez szükséges információt a kapcsolatokról még minden csak a BGP routing táblákból szereztek. Volt azonban egy újfajta megközelítést is, ami az önálló AS-kapcsolatokat tárta fel az Internet router-szintű topológiájából [4]. Ez több szempontból is előnyös volt, mint a BGP-s routing táblákból kiolvasott topológia, hiszen

1. Nagyobb felbontásban látjuk az AS-szintű térképet (pl. látszódnak a többszörös kapcsolatok AS-ek között);
2. Látszódnak a BGP protokoll által aggregált - így az AS-szintű hálózatban eltakart - útvonalak;
3. Így már lehetőség volt azonosítani a határ-routereket⁴, aminek segítségével pontosabban karakterizálhatták az AS-en belül kapcsolatokat.

Az addigi eredményeket felhasználva már volt egy viszonylag pontos kép az Internet tartomány-szintű topológiájáról, amit tovább pontosítottak úgy, hogy az egyes AS-eken belüli szabályokat is feltárták, hiszen keveset tudtak még arról, hogy milyen routing policy-t használnak az AS-k. A BGP protokoll ugyanis lehetővé teszi az AS-eknek, hogy megválasszák az útvonalválasztási policy-jukat, ami alapján történik a csomagtovábbítás és az elérhetőségi adatok terjesztése az AS-en belül. Megmutatták, hogy az AS-k a többi szolgáltatók csak egy csoportjának hirdetik magukat, ami mögött valószínűleg traffic engineering⁵ lehet. Pl. több Tier-1-es⁶ AS is az ügyfeleit (közvetve vagy direkt) a peer kapcsolatain keresztül éri el a közvetlen customerei helyett. Ezenkívül a válogatott hirdetés szerint jóval kevesebb elérhető útvonal van az Interneten, mint azt az AS kapcsolati gráf mutatja (Ezért is volt jóval pontatlanabb a BGP routing táblák alapján elképzelt kép.) [5].

Miután már volt valamilyen fogalmunk arról, hogy az egyes AS-k hogyan működtetik hálózatukat, az AS-ek közötti kapcsolatok mélyebb megismerése következett. Adva van az Internetnek már megkülönböztetett router-szintű és AS-szintű topológiái, illetve felületesen már ismerjük az egyes policy-kat. Azt kezdték el vizsgálni, hogy a router-szintű topológiában mi lenne a legrövidebb út és ehhez képest mi az adott (valós) policy-út [6]. Több szempontból is vizsgálták a kérdést:

⁴Azon routereket, amelyek az AS szélén vannak és a többi AS-hez jelentik a kapcsolatot, határ-routernek nevezzük.

⁵Forgalomszabályozás: tudatos tervezéssel próbálják meg elkerülni azokat az eseteket, amikor szolgáltatás-kimaradás lép fel a túlterhelés miatt.

⁶A BGP hierarchiában a legmagasabb szintű szolgáltatók, ezeket követi a Tier-2 és a Tier-3.

- Mennyivel fújja fel a policy a shortest-utat⁷?
- Létezik-e S és D forrás-cél pár között egy I köztes pont, amire a $d_{policy}(S \rightarrow I) + d_{policy}(I \rightarrow D) < d_{policy}(S \rightarrow D)$? (Ebben az esetben lehetne egy ilyen közbeiktatott ponttal javítani)
- A policy-k vajon a nagyobb AS-k felé terelik a forgalmat?

Ugyan abban az időben még csak a legrövidebb utakat vizsgálták és azokat is a multicasting⁸ szempontjából, de azóta a kutatók felismerték, hogy olyan kérdések várnak még válaszra, amelyek alapvetően befolyásolják az egész Internet gerinchálózatát. Ezenkívül jelentősen megkönnyítené, illetve pontosabbá tehetné a téma körbeli kutatásokat, ha rendelkezésünkre állna egy olyan hálózati modell, amely nem csak a skálafüggetlenséget⁹ tudja szimulálni, hanem a - gazdasági érdekek miatt titkolt - kapcsolatokat és útvonalválasztást is valóságban tudja modellezni.

Közeli húsz év óta foglalkoztatja a kutatókat a hálózatok kialakulása és a kialakult útvonalak, ugyanakkor az Internet AS-szintű topológiának vizsgálatakor arról nem állítanak semmit, hogy az egyes policy-kat miért használják. Sőt, ugyanígy nem tudunk szinte semmit a már említett más területekről származó problémák esetében sem.

Jelen Diplomaterv keretében egy olyan megoldást mutatok be, amely képes megbecsülni, hogy miért - milyen szabályszerűségek figyelembevételével - hoztak meg egyes döntéseket. Azzal a feltételezéssel éllek, hogy minden policy-t fel lehet építeni jól meghatározott, egyszerű primitívekből, illetve ezen primitívek azonosítása és változatos összekapcsolása révén meg fogjuk tudni mondani, hogy egy már kialakult hálózatot milyen komplex policy határoz meg.

1.2. Útvonalválasztás, routing algebrák

Az általános hálózati modellben egy véges, egyszerű és összefüggő $G(V; E)$ gráfot használunk a hálózat megadására, ahol $|V| = n$ és $|E| = m$. Az élek irányítottsága és súlyozása az adott problémától függ, minden kombináció megengedett. Legyen $\deg(v)$ a $v \in V$ csúcs fokszáma és legyen $d = \max_{v \in V} \deg(v)$. Egy $s - t$ séta a csúcsoknak egy $p = (s) = v_1, v_2, \dots, v_k (= t)$ sorozata, ahol k a séta hossza és $(v_i, v_{i+1}) \in E : \forall i = 1, \dots, k-1$. Ezenkívül egy kör olyan séta, ahol $s = t$.

Ahhoz, hogy meg tudjuk mondani, hogy két útból melyiket érdemes választani, definiálni kell egy preferencia sorrendet az utak között. Ehhez pedig az utak hosszára lesz

⁷Shortest-út: Azt a policy-t, amely a legrövidebb utakat választja, shortest-path (legrövidebb-út) policy-nak nevezzük.

⁸A multicast egy információtovábbítási mód, amikor egy üzenetet több célhoz szeretnénk eljuttatni egyidejűleg egyetlen forrástól.

⁹Egy hálózat skálafüggetlen, ha benne a fokszámeloszlás hatványfüggvényt követ. Az Internet hálózata tipikusan skálafüggetlen hálózat. (A kifejezés Barabási Albert László magyar származású amerikai matematikus nevéhez fűződik.)

szükség, amit egy metrikával, azaz távolságfüggvényel tudunk mérni. A metrikus tér fogalma egy (halmaz, függvény) párt jelent, ahol a függvény bármely két, halmazbeli elemhez egy nemnegatív valós számot rendel (vagyis a távolságukat méri).

1.2.1. Definíció (Metrikus tér). *A metrikus tér egy olyan (X, δ) pár, ahol X egy tetszőleges halmaz,*

$\delta: X^2 \rightarrow \mathbb{R}_0^+$ pedig egy olyan nemnegatív valós függvény, melyre tetszőleges $x, y, z \in X$ esetén:

1. $\delta(x, y) = 0 \Leftrightarrow x = y$ (megegyezőségi tulajdonság)
2. $\delta(x, y) = \delta(y, x)$ (szimmetria)
3. $\delta(x, z) \leq \delta(x, y) + \delta(y, z)$ (háromszög-egyenlőtlenség).

Annak érdekében, hogy minél általánosabb lehessen a modell, minden $e \in E$ élet egy tulajdonság-vektor jellemz, amely vektor minden dimenziójának értékeit különböző metrikák szerint adhatunk meg. Így definiálhatjuk az él hosszát, (sáv)szélességét, késleltetését, megbízhatóságát, sőt bármilyen nem tipikus tulajdonságot is (pl. szín, vagy egy időtől függő $f(t)$ függvényt, stb.).

A metrikák fontos szerepet játszanak az útvonalválasztásban, hiszen ez alapján tudjuk a hálózati utak közötti preferenciát megadni, másnépp megfogalmazva megadja, hogy egy bizonyos tulajdonság alapján mekkora a költsége egy útnak. Az útvonalválasztás során a metrika lehet statikus, amikor egy előre rögzített elvet követünk végig, vagy lehet dinamikus, amikor a hálózat adott állapotától függően automatikusan változik. Azt, hogy milyen metrika szerint végezzük a routing-ot nevezzük *routing policy*-nak:

1.2.2. Definíció (Routing policy). *A routing policy egy olyan $p_{st}^* = Pol(\mathcal{P}_{st})$ függvény, aminek az értelmezési tartománya a lehetséges $s-t$ utak: \mathcal{P}_{st} és az adott policy-nak megfelelő legkedvezőbb utat adja vissza.*

Ahhoz, hogy ezentúl matematikailag is kényelmesen tudjuk kezelni a policy-kat, az ún. routing algebrák fogalmát kell használnunk, melyek az általános útvonalválasztó policy-k matematikai leírása [7, 8].

1.2.3. Definíció (Routing algebra). *Az \mathcal{A} routing algebra¹⁰ egy teljesen rendezett félcsoport egy „végétlen elemmel”: $\mathcal{A} = (W, \phi, \oplus, \preceq)$, ahol W az élek súlyainak halmaza, ϕ ($\phi \notin W$) egy speciális végtelen súly, abban az értelemben, hogy azon az élen vagy úton nem lehet átmenni, a \oplus a súlyok egy kétváltozós kompozíció operátora, a \preceq pedig a súlyok összehasonlító operátora.*

Még pontosabban, a következő tulajdonságokat követeljük meg:

- (W, \oplus) egy Abel-csoport:

¹⁰Az algebra elnevezés arra utal, hogy - mint később látni fogjuk - műveleteket lehet végezni ezen objektumokon.

- *Zárt:* $w_1 \oplus w_2 \in W, \forall w_1, w_2 \in W$
- *Asszociatív:* $(w_1 \oplus w_2) \oplus w_3 = w_1 \oplus (w_2 \oplus w_3), \forall w_1, w_2, w_3 \in W$
- *Kommutatív:* $w_1 \oplus w_2 = w_2 \oplus w_1, \forall w_1, w_2 \in W$
- \preceq teljes rendezés W -n:
 - *Reflexív:* $w \preceq w, \forall w \in W$
 - *Anti-szimmetrikus:* Ha $w_1 \preceq w_2$ és $w_2 \preceq w_1$, akkor $w_1 = w_2, \forall w_1, w_2 \in W$
 - *Tranzitív:* Ha $w_1 \preceq w_2$ és $w_2 \preceq w_3$, akkor $w_1 \preceq w_3 \forall w_1, w_2, w_3 \in W$
 - *Teljes:* $\forall w_1, w_2 \in W: w_1 \preceq w_2$ vagy $w_2 \preceq w_1$
- ϕ összeegyeztethető a (W, \oplus) Abel-csoporttal \preceq szerint:
 - *Elnyelés:* $w \oplus \phi = \phi, \forall w \in W$
 - *Maximalitás:* $\phi \not\preceq w, \forall w \in W$

1.2.1. Megjegyzés. Kiemelném az összehasonlítás operátor (\preceq) teljes rendezési tulajdon-ságát, mert az a tulajdonság teszi lineárissá¹¹ a rendezést.

1.2.2. Megjegyzés. A ϕ végtelen elem pusztá létét, ill. összeegyeztethetőségét az Abel-csoporttal szintén fontos kiemelni, hiszen ezáltal szinten bármilyen alaphalmazt megadhatunk az élek bármelyik, tulajdonság-vektorbeli dimenziójának. Emellett érdemes megjegyezni, hogy azon (részben)rendezett halmazok, melyek bármely kételemű részhalmazának létezik infimuma és szuprimuma, hálóknak nevezük. A routing algebrák esetében az természetesen csak akkor teljesül, ha alkalmasan választjuk meg a rendező (\preceq) operátorát és az alaphalmazt: a valós számhalmazon, a „hagyományos” \leq rendezés esetén a routing algebrák hálók.

Most már meg tudjuk mondani egy egyszerű él súlyát. Az útvonalválasztás során azonban nem éleket akarunk összehasonlítani, hanem útvonal-alternatívákat:

1.2.4. Definíció (Egy út súlya). Egy $p = (v_1, v_2, \dots, v_k)$ út $w(p)$ hosszát az út éleinek súlyainak \oplus -szerinti összege adja:

$$w(p) = \bigoplus_{i=1}^{k-1} w(v_i, v_{i+1}).$$

Ezután azt mondjuk, hogy egy preferált (tetsző, legjobb, stb.) út az \mathcal{A} algebrában s és t között a legkisebb súlyú \preceq szerint:

$$Pol(\mathcal{P}_{st}) = p^* : w(p^*) \preceq w(p), \forall p \in \mathcal{P}_{st}.$$

Ezek után könnyen ellenőrizhető, hogy a legáltalánosabban használt routing policy, a shortest path routing (legrövidebb utak) algebrája a $(\mathbb{R}^+, \infty, +, \leq)$, míg egy másik policy, a widest-path routing (legszélesebb utak) algebrája a $(\mathbb{R}^+, 0, \min, \geq)$.

¹¹Az első három tulajdonság miatt csak részbenrendezett halmazról beszélhetünk, ha azonban minden elem összehasonlítható, akkor válik teljes, vagy lineáris rendezéssé \preceq .

1.3. Routing algebrák tulajdonságai

A routing algebráknak, mint matematikai struktúrának számos érdekes tulajdonsága van. Ezek közül vannak olyanok, melyek alapvetően befolyásolják az algebra felhasználhatóságát, hiszen az algebra olyan minőségbeli tulajdonságát határozzák meg, mint például az útvonalválasztás algoritmikus lépésszáma. Vannak pusztán leíró jellegű tulajdonságok is, melyek segítenek az algebrák összehasonlításában, az egyes policy-primitívek kiválasztásában is.

1.3.1. Monotonitás, izotonitás és egyéb tulajdonságok

A legtöbb esetben két meghatározó tulajdonsággal kell rendelkeznie egy algebrának, hogy „jól viselkedőnek” mondjuk. Az ilyen algebrákat reguláris algebrának nevezzük:

1.3.1. Definíció (Reguláris algebra). Az \mathcal{A} routing algebrát regulárisnak nevezzük, ha

- Monoton (M): $w_1 \preceq w_1 \oplus w_2, \forall w_1, w_2 \in W$
- Izotón (I): $w_1 \preceq w_2 \Rightarrow w_3 \oplus w_1 \preceq w_3 \oplus w_2, \forall w_1, w_2, w_3 \in W$

A monotonitás (M) azt jelenti, hogy egy w_1 súlyú élet, elé illesztve egy w_2 súlyú másik éssel csak kevésbé preferáltható teheti: $w_1 \preceq w_2 \oplus w_1$. A kommutativitás miatt igaz az él után való illesztésre is: $w_1 \preceq w_1 \oplus w_2$. A hagyományos értelemben vett hosszúságot általánosítja ez a tulajdonság, azaz a „a hosszabb út rosszabb”¹².

Az izotonitás (I) azt jelenti, hogy adott \preceq relációban álló éleket ugyanazzal az éssel előlről, vagy hátulról meghosszabbítunk, akkor a relációs viszony nem változik (Az itt leírtak igazak élek helyett utakra is).

Az F.2. függelékben megtalálható a routing teljes modellje, melyet a routing algebrák tervezésekor használtak [9]. Ez a modell a lokális memóriaigényre koncentrál, és azt mondja, hogy egy \mathcal{A} routing algebra tömöríthetetlen, ha az $\mathcal{M}_{\mathcal{A}}$ lokális memóriaigény $\Omega(n)$, különben \mathcal{A} tömöríthető. Egy tömöríthetetlen routing algebra nyilván nem skálázódik jól, viszont a tömöríthető algebrák igen. A reguláris algebrák „jól viselkednek”, hisz a monotonitás és izotonitás garantálja, hogy a preferált út megkapható polinom időben az általánosított Dijkstra algoritmussal. Ez lehetővé teszi, hogy legfeljebb $\tilde{O}(n)$ bit információt tároljunk lokálisan, így nem csak az elméletben, hanem a valóságban is használható algebrákat kapunk [9, 10].

A teljesség igénye nélkül felsorolok még néhány tulajdonságot [11]:

- *Delimitált (D):* $w_1 \oplus w_2 \neq \phi, \forall w_1, w_2 \in W$
- *Szigorúan monoton (SM):* $w_1 \prec w_1 \oplus w_2, \forall w_1, w_2 \in W$
- *Kiválasztó (S):* $w_1 \oplus w_2 \in \{w_1, w_2\}, \forall w_1, w_2 \in W$

¹²Pontosabban a hosszabb út nem jobb, mint a rövidebb.

- *Elnyelő (C):* $w_1 \oplus w_2 = w_1 \oplus w_3, \forall w_1, w_2, w_3 \in W$

A fentiek közül talán csak a delimitáltság (D) szorul magyarázatra. Ez a tulajdonság garantálja, hogy az éleket bármilyen önkényesen választott sorrendben is kombinálva járható utat kapunk. Általában az intra-domain¹³ routing policy-k rendelkeznek ezzel a tulajdon-sággal, de vegyük csak példának a BGP völgymentességét: ha egy csomag a hierarchiában fentebbi AS-től érkezik, akkor azt már csak lefelé mutató, vagy peer kapcsolaton keresztül lehet továbbítani.

1.4. Műveletek routing algebrák között

A routing algebrák megadási módja igen változatos policy-k leírására ad lehetőséget, de - annak érdekében, hogy teljes rendezés lehessen \preceq - minden csak egy metrika szerint tudunk optimalizálni. Ha szeretnénk több szempontot is figyelembe venni, amire van is példa a routing policy-k között, akkor ezt az algebrák egymásutáni alkalmazásával tudjuk megtenni, így néhány egyszerű policy algebrájával meglepően bonyolult és a valóságot nagyon jól leíró algebrákat tudunk létrehozni [10]. Két ilyen műveletet igen fontos szerepet kap, az egyik a lexikografikus szorzat, ami az összeillesztő-, és a subalgebra, ami a szétválasztó operátor [11].

1.4.1. Definíció ($\mathcal{A} \times \mathcal{B}$). Legyenek adottak az $\mathcal{A} = (W_{\mathcal{A}}, \phi_{\mathcal{A}}, \oplus_{\mathcal{A}}, \preceq_{\mathcal{A}})$ és $\mathcal{B} = (W_{\mathcal{B}}, \phi_{\mathcal{B}}, \oplus_{\mathcal{B}}, \preceq_{\mathcal{B}})$ algebrák. Ekkor az $\mathcal{A} \times \mathcal{B} = (W, \phi, \oplus, \preceq)$ lexikografikus szorzatuk, ahol

- $W = W_{\mathcal{A}} \times W_{\mathcal{B}}$, ill. $\phi = (\phi_{\mathcal{A}}, \phi_{\mathcal{B}})$
- $(w_1, v_1) \oplus (w_2, v_2) = (w_1 \oplus_{\mathcal{A}} w_2, v_1 \oplus_{\mathcal{B}} v_2), \forall w_1, w_2 \in W_{\mathcal{A}}$ és $v_1, v_2 \in W_{\mathcal{B}}$
- $(w_1, v_1) \preceq (w_2, v_2) = \begin{cases} v_1 \preceq_{\mathcal{B}} v_2 & \text{ha } w_1 =_{\mathcal{A}} w_2 \\ w_1 \preceq_{\mathcal{A}} w_2 & \text{különben} \end{cases}$

1.4.1. Megjegyzés. A ϕ jól definiált, ha \mathcal{A} és \mathcal{B} delimitáltak. Egyéb esetben ϕ megadása nem magától értetődő.

1.4.2. Megjegyzés. A lexikografikus szorzat nem kommutatív: $\mathcal{A} \times \mathcal{B} \neq \mathcal{B} \times \mathcal{A}$.

A korábban már említett shortest-path és a widest-path policy-k algebráinak lexikografikus szorzata az \mathcal{SW} (shortest-widest - legrövidebb-legszélesebb routing: a legszélesebb útra irányít, ám ha több ilyen is van, akkor azok közül a legrövidebben) és a \mathcal{WS} (widest-shortest - legszélesebb-legrövidebb routing: először a legrövidebb útra irányít, ha több ilyen is van, akkor azok közül a legszélesebbben).

A második definiált művelet a subalgebra:

1.4.2. Definíció ($\hat{\mathcal{A}}$). Adott az $\mathcal{A} = (W, \phi, \oplus, \preceq)$ algebra és a súlyok egy $\hat{W} \subseteq W$ részhalmaza. Az \mathcal{A} algebra leszűkítése \hat{W} -re: $\hat{\mathcal{A}} = (\hat{W}, \phi, \oplus, \preceq)$ akkor és csak akkor subalgebrája \mathcal{A} -nak, ha \hat{W} zárt \oplus -ra.

¹³Intra-domain: AS-n belüli.

1.4.1. A műveletek hatása a tulajdonságokra

A routing algebrák kompozíciójaként létrejött új algebrák tulajdonságai levezethetők az alkotó algebrákéból.

A következő tétel mutatja a lexikografikus szorzat hatását a tulajdonságokra:

1.4.1. Tétel (A lexikografikus szorzat hatása az algebrák tulajdonságaira.). .

- $M(\mathcal{A} \times \mathcal{B}) \Leftrightarrow SM(\mathcal{A}) \vee (M(\mathcal{A}) \wedge M(\mathcal{B}))$
- $I(\mathcal{A} \times \mathcal{B}) \Leftrightarrow I(\mathcal{A}) \wedge I(\mathcal{B}) \wedge (N(\mathcal{A}) \vee C(\mathcal{B}))$
- $SM(\mathcal{A} \times \mathcal{B}) \Leftrightarrow SM(\mathcal{A}) \vee (M(\mathcal{A}) \wedge SM(\mathcal{B}))$

Az (1.4.1) tétel szerint annak, hogy reguláris algebrákat hozunk létre, szükséges feltétele, hogy csakis izonton algebrákat használunk fel, sőt a monotonitás is megkövetelt minden két tagra vagy az első tag szigorú monotonitása.

1.5. Példák algebráakra és kombinálásukra

Az 1.5.1 táblázatban néhány példát látunk a leginkább kutatott intra-domain routing policy-kra az algebráikkal és a legfontosabb tulajdonságaikkal. Az utolsót leszámítva az összes felsorolt algebra delimitált és reguláris (D, M, I).

1.5.1. táblázat. *Routing policy-k, algebráik és fontosabb tulajdonságaik.*

Policy	Algebra	Tulajdonság
Legrövidebb út	$\mathcal{S} = (\mathbb{R}^+, \infty, +, \leq)$	SM, I
Legszélesebb út	$\mathcal{W} = (\mathbb{R}^+, \infty, \min, \geq)$	S, I, M
Legmegbízhatóbb út	$\mathcal{R} = ((0, 1], 0, *, \geq)$	SM, I
Legszélesebb-legrövidebb út	$\mathcal{WS} = \mathcal{S} \times \mathcal{W}$	SM, I
Legrövidebb-legszélesebb út	$\mathcal{SW} = \mathcal{W} \times \mathcal{S}$	SM, $\neg I$

A \mathcal{W} a widest-path routing algebrát jelenti [12]. Ennél az algebránál egy él súlya a kapacitása adja, és nyilván egy több élből álló úton az út kapacitása megegyezik az útmenti legkisebb kapacitással. Emellett nyilván minél nagyobb egy út globális kapacitása, annál inkább preferált. A \mathcal{W} algebra kiválasztó is és megadható a $(\mathbb{R}^+, \infty, \min, \geq)$ négyessel (az (F.3.1) tétel értelmében pedig tömöríthető is).

A legmegbízhatóbb út algebra (\mathcal{R}) úgy értelmezhetjük, hogy az élek súlyát az a valószínűség adja, hogy az adott élen a csomag hibátlanul át tud menni, így nagyobb érték a jobb. Nyilvánvaló, hogy \mathcal{R} tartalmaz egy szigorúan monoton (SM) subalgebrát: ha nem engedjük meg a biztos valószínűséget, akkor minden út biztosan romlik, ha hozzáveszünk egy újabb élet, azaz az értéke csökken, mert egy 1-nél kisebb számmal szorozzuk az eddig súlyt: $\hat{\mathcal{R}} = ((0, 1], 0, *, \geq)$.

A legszélesebb-legrövidebb út (widest-shortest path) \mathcal{WS} routing esetén a legrövidebb utak közül a legnagyobb kapacitású választjuk [13], míg a legrövidebb-legszélesebb út (shortest-widest path) \mathcal{SW} ([12, 14]) routing esetén, éppen fordítva, a legnagyobb kapacitású utak közül a legrövidebbet választjuk. Ezeket az algebrákat megkaphatjuk a \mathcal{S} és a \mathcal{W} algebrák lexikografikus szorzataként [11].

1.5.1. Megjegyzés. \mathcal{SW} nem izotón. Az (F.3.2) tételel áll a nem izotón algebrákra is, így $\Omega(n)$ bit lokális memóriát igényel a \mathcal{SW} algebra is. Jelenleg még nyitott kérdés, hogy ez a határ szoros-e. Nem tudjuk, hogy van-e jobb megoldás, mint a triviális, azaz hogy minden router tárol egy-egy routing tábla bejegyzést minden forrás-cél párra, ami $O(n^2 \log d)$ (d a maximális fokszám) bit per router memóriaigényű.

1.6. Összefoglaló

Ebben a fejezetben áttekintettem a szakirodalmat, összeszedve a legfontosabb állomásokat. Az 1990-es évek végétől egyre többet foglalkoztak a kutatók a BGP hálózatának feltárással, amivel el is jutottak addig, hogy van egy viszonylag pontos kép a hálózatról, de arra, hogy miért egy adott policy-t használnak az AS-ek, nem tudnak válaszolni. Rámutattam, hogy az Internet AS-szintű topológiáján kívül, a más tudományterületekről származó problémák útvonalválasztásáról sem tudunk sok minden és ezért van szükség egy olyan eszközre, ami a policy-feltáras feladatát - általános esetben is - hatékonyan el tudja láttni. Ehhez definiáltam a routing algebrát, bemutattam a legfontosabb tulajdonságait és műveleteit, emellett a legszélesebb körben használt policy-k algebráit is ismertettem.

2. fejezet

A hálózatkutatás legfontosabb modelljei

Számos kutatás foglalkozik különböző hálózatokon terjedő információk vizsgálatával. Természetesen minden ilyen modellt le lehet írni az általanosított hálózati modellel, így a meghatározó policy-k azonosítása után vizsgálható az 1. fejezetben bemutatott, vagy azokhoz hasonló algebrákkal. Ebben a fejezetben bemutatom a hálózatkutatás leggyakrabban vizsgált modelljeit, és leírom a modellekhez tartozó meghatározó policy-ket is.

Az útvonalválasztás szempontjából alapvető különbség van a lokális- és a globális optimalizálás között. Míg globális optimalizálás esetén az egész hálózatot figyelembe véve választjuk meg a legjobb útvonalat, addig lokális optimalizálásnál csak a helyi viszonyok számítanak. Jelenleg a számítógépes hálózatokban a két legelterjedtebben használt policy kategória a „distance-vector”¹ és a „link-state”². A módszerek abban különböznek, hogy a hálózatról milyen információkat gyűjtenek és hogyan, de abban megegyeznek, hogy globális optimumra törekszenek, tehát pl. a legrövidebb utat keresik meg. Ezzel szemben pl. a greedy routing³ csupán törekszik elérni globális optimumot, ám könnyen elakadhat, ha nem javítunk az alapalgoritmuson [15]. Könnyen belátható, hogy a vírusterjedés-jellegű feladatok lokális optimumra törekszenek, míg az Internetes útvonalválasztás globálisan optimalizál.

A másik szempont, ami meghatározó, az az, hogy a csomópontok felülről tudják-e bírálni a csomagok terjedését abban az értelemben, hogy a küldő eredeti célját meg tudja-e valósítani akkor is, ha az útvonal mentén valamelyik csomópont ezt nem akarja. Feltehetjük-e, hogy a hibamentes csomagtovábbítás elsődleges prioritás minden hálózati csomópontnak (pl. kölcsönös bizalom megteremtése érdekében), vagy az egyes csomópontok viselkedhetnek önkényesen?

Mivel ezen két tulajdonság nem zárja ki egymást, a legtöbbször azzal az esettel állunk szemben, hogy nem tudunk egy problémát egyértelműen karakterizálni.

¹Távolság-vektor módszerek, a Bellman-Ford algoritmusra épül.

²Link-állapot módszerek, a Dijkstra algoritmusra épül.

³Mohó útvonalválasztás: A lokális helyzet optimumát választja.

2.1. Vírusterjedés komplex hálózatokban

A vírusok terjedését és járvánnyá fejlődését sok évtizede vizsgálják. A járványok előrejelzése lehetőséget ad a tudósoknak, hogy megtervezzék a védőoltások ütemezését és az esetleges karanténok felállítását, ami jelentős hatással lehet egy adott betegség halálozási rátájára. A fertőző betegségek modellezése egy olyan eszköz, ami segítséget ad a fertőzés terjedésének vizsgálatához és a jövőbeli járványok elkerüléséhez stratégiákat lehet kidolgozni általa [16].

2.1.1. A vírusterjedés matematikai modellje

A legkorábbi ismert matematikai modell a fertőző betegségek terjedéséről Bernoullitól származik, aki a fekete himlőt vizsgálva megállapította, hogy ha mindenkit beoltanának a betegség ellen, akkor az átlagos életkor 26 év 7 hónapról 29 év 9 hónapra nőne [17].

Bernoulli munkája során természetesen még nem értette annyira a baktériumok és vírusok biolóját, mint mi. A XX. század első felében Ronald Ross malária kutatásával kezdetét vette a modern elméleti fertőzés kutatás. Ezután nem sokkal, 1927-ben A. G. McKendrick and W. O. Kermack elismert munkája is megjelent, az „A Contribution to the Mathematical Theory of Epidemics”. Ez a determinisztikus modell sikeresen jelezte előre sok fertőző betegség viselkedését.

Amikor nagy populációkat vizsgálunk, a determinisztikus modelleket használjuk, mert egy ilyenben a populáció egyedeit besorolhatjuk alcsoportokba, ami a fertőzés egy specifikus stádiumát jelenti. Ha a fertőzés átterjedése az egyik csoportról a másikra időben folytonos, a csoportok aktuális mérete matematikailag a deriválással fejezhető ki, így a modell leírható differenciál egyenletekkel. Ott játszik fontos szerepet a determinisztikusság, hogy egy ilyen modellben feltesszük, hogy egy csoport egyedszáma deriválható az idő szerint, ehhez viszont az kell, hogy a fertőződés kiszámíthatóan (nem véletlenszerűen) terjedjen tovább, más szóval a csoportok egyedszámának változása egyértelműen meghatározható az addigi múltból. Ezen modellek szakirodalmát és terminológiáját áttanulmányozva számos különböző mérőszámot és tulajdonságot találhatunk, de az alapmodell a következő három csoportot jelöli:

- $S(t)$: A t időpontig meg nem fertőzött egyedek száma, azaz akik még megfertőződhetnek.
- $I(t)$: A t időpontig megfertőzött egyedek száma, azaz akik tovább tudnak fertőzni.
- $R(t)$: A t időpontig megfertőzött, de mégsem $I(t)$ -beli egyedek száma, azaz vagy meggyógyultak, vagy meghaltak. Ezen csoport egyedei már nem tudnak újra megbegegni, sőt fertőzni sem fertőznek.

Egy egyed az $S(t) \rightarrow I(t) \rightarrow R(t)$ sorrendben halad a csoportok között. Egy fix méretű

populációt feltételezve, $N = S + I + R$, a következő egyenleteket vezethetjük be [18]:

$$\frac{dS}{dt} = -\beta SI \quad (2.1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (2.2)$$

$$\frac{dR}{dt} = \gamma I \quad (2.3)$$

Ahol β a kapcsolati ráta és $1/\gamma$ az átlagos fertőző periódus.

Néhány feltételezést tettünk az egyenletek felírásához: (1) minden egyedet ugyanakkora valószínűséggel fertőz meg egy β kapcsolati rátájú betegség, így minden fertőzött egyed βN egészséges egyedet fertőz meg egy időegység alatt, illetve az egyedeknek átlagosan S/N megfertőzhető kapcsolata van. Tehát a fertőzási ráta: az új betegek száma $\beta N(S/N)$, azaz amennyivel változik (nyilván csökken) S : $\beta N(S/N)I = \beta SI$.

Ezután érthető a második és harmadik egyenlet: a fertőzöttek száma annyival nő, ahánnyal kevesebb egészséges lesz, illetve annyival fogy, ahány meghal vagy meggyógyul. A γ jelöli az átlagos halálozási/felépülési rátát.

Végül feltessük, hogy a megfertőződés és a felépülés időben jóval gyorsabb folyamat, mint a születés és halálozás, így ezeket a faktorokat kihagyhatjuk a modellből.

2.1.1. Megjegyzés. Természetesen számos következő szempontot figyelembe lehet még venni a modell mélyítése érdekében, ilyen pl. az előbb említett születés és halálozás; az $R \rightarrow S$ átmenet, azaz a fertőzésből való meggyógyulás után újra meg lehet betegedni; figyelembe lehet venni az ún. lappangási időszakot, bevezetve egy újabb csoportot ($S \rightarrow \epsilon \rightarrow I \rightarrow R$); vagy a veleszületett betegséget is, azaz a fertőzött anyától elkapott betegséget is.

Jelen dolgozat szempontjából az alap SIR modell megfelelő.

2.1.2. Vírusterjedés, mint útvonalválasztási probléma

A fertőző betegségek terjedése nem látszik útvonalválasztási kérdésnek sőt, semmilyen policy-t nem veszünk észre első ránézésre, ezért talán némi magyarázatot igényel a problémakör vizsgálata. A diplomadolgozat végső célja rejtett szabályok felismerése előre definiált policy primitívek segítségével tetszőleges hálózatban, így világos, hogy a vírusterjedést sem zárhatom ki a vizsgálatból.

A fejezet bevezetőjében említett két tulajdonság közül az optimalizálás az, ami itt a legszembetűnőbb. Ha ugyanis nem úgy tekintünk a problémára, mint egy kétszereplős „háború” (ember a vírus ellen), hanem úgy, mint a kis organizmusok egyéni útvonalválasztására, akkor látható, hogy lokális optimalizálásról van szó, hiszen minden egyes fertőző elem szeretne minél tovább élni újabb és újabb gazdatestet találva magának. Lévében itt nem a tipikus több lépésen keresztüli pont-pont routing-ról van szó, globális optimalizálási feladat legfeljebb az lehet, hogy minél több csomópont legyen fertőzött. minden csomópontból arra fertőz tovább a vírus, amerre a legtöbb esélytel fog továbbélni - ez nyilván függ az adott vírus-egyedtől is -, azaz pl. a leggyengébb immunrendszerű ember felé. Ez

alapján két policy-t mutatok be:

- Fertőzési-határ: A vírus-egyed legjobb továbbfertőzési policy-ja.
- Unió-fedés: A globális fertőzés policy-ja.

Fertőzési-határ

A Fertőzési-határ policy alapú útvonalválasztás garantálja a legjobb túlélési esélyt egy vírus-egyed számára. Ennek a policy-nak az algebrája az $\mathcal{F} ((0, 1], 0, \max, \geq)$. Ennél az algebránál az élek súlya (egy $(0, 1]$ intervallumbeli valós szám) azt adja meg, hogy azt a csomópontot, ahol vezet, milyen valószínűsséggel tudjuk megfertőzni, így nyilván a nagyobb jobb, a 0 pedig azt jelenti, hogy a csomópont $R(t)$ -beli, azaz már biztosan nem lehet megfertőzni. Mivel az útvonalválasztás lokálisan optimalizál, ezért két alternatív élet a \geq operátorral tudunk összehasonlítani, illetve a két él összeillesztésére a két él súlyának maximumát vesszük.

Unió-fedés

Az Unió-fedés policy szerinti útvonalválasztás célja, hogy lehetőleg olyan élen menjünk és fertőzzük tovább, amit még nem használtak. Ennek a policy-nak a számítógépes vírusok tárgyalásánál van nagyobb szerepe, semmint a biológiai fertőzéseknel, hiszen ez alapján - akár már fertőzött gépet is - minden olyan élet saját felügyelet alá vehetünk, amit még nem találtunk meg, így például ügyelhetünk arra, hogy semelyik fertőzött gép ne tudja frissíteni a vírusirtójának a vírusdefiníciós adatbázisát.

Az Unió-fedés policy algebrája az $\mathcal{U} (\mathbb{N}, \infty, f, \leq)$. Itt az élsúly egy természetes szám lehet és azt fejezi ki, hogy hányszor használták már fel egy útvonalválasztásban. Mivel a cél az, hogy minél ritkábban (optimálisan soha nem) használt utakat használunk, a ∞ jelenti az átjárhatatlanságot és a \leq az összehasonlító operátor. A több elől álló út súlyát az f összegző függvény adja:

$$f(e_1, e_2) = \begin{cases} e_1 + e_2 & \text{ha } e_1 * e_2 \neq 0 \\ 0 & \text{különben} \end{cases}$$

Ezzel garantáljuk, hogy olyan esetben, amikor úgyis mindegy - azaz nincs még fel nem használt él - akkor az élek súlyának összegét vesszük, de amikor van olyan él, amit még nem használtunk, akkor azt biztosan fogjuk használni (hacsak nincs egy másik alternatív 0 súlyú út), hiszen a két élű út súlya 0 lesz.

2.2. Trendterjedés közösségi hálózatokban

A közösségi hálózatok az elmúlt 10 év egyik legmeghatározóbb jelensége, új korszakot nyitott az emberi kapcsolatokban. A kapcsolatháló elemzés a szociometriából indult, de annál bonyolultabb, időben is változó, nagyobb közösségek vizsgálatára alkalmas módszer-tan. A kapcsolatháló elemzési módszerek alkalmasak egy nagyobb csoport klickjeinek és alcsoporthajnak felrajzolására és megjelenítésére is. A kapcsolathálózati megközelítés kevés-bé hangsúlyozza az egyéni cselekvés szerepét a struktúrák létrehozásában, nem az egyéni szándékból, hanem a struktúrák belső feszültségeiből jön létre a cselekvés mozgástere. A struktúra - pl. egy elterjed trend - e szemléletmód számára nem egy közvetlenül megmutatkozó adottság, hanem sokkal inkább a kapcsolatok hálójából bonyolultan kibontható szerveződés.

A hálózati megközelítés előnye, hogy az adott hálózati struktúrákat, közösségeken belül kialakuló attribútumokat dinamikus módon, a változásokon keresztül is képes vizsgálni: a kapcsolatok és a környezet folytonosan változó mozgása mentén. A hálózat elemzéssel komplex módon tudjuk leírni vizsgált közösségi hálózatok működését [19].

A trend- vagy divatterjedés egy szociális hálóban lehetséges, ahol az egyik csomóponttal szimbolizált egyén felvesz egy szokást, elkezd hordani egy bizonyos ruhadarabot (vagy márkát), vagy bármi olyat tesz, amit „az átlag” nem. Ha az ismerősei - a kapcsolati hálóban szomszédos csomópontok - észreveszik ezt az újdonságot és megtetszik nekik is, átveszik a forrástól. A problémakör alapvető kérdése, hogy mi határozza meg, hogy elterjed-e egy trend és divattá válik-e az egész hálózatban, vagy sem. Malcolm T. Gladwell brit-kanadai újságíró, író 2000-ben megjelent könyvében [20] számos példát olvashatunk, hogy egyes szituációkban mikor érkezett el a fordulópont, mikor vált általános divattá valami. Ilyen példa a Hush Puppy cipőgyártó cég, mely egyik pillanatról a másikra a csőd széléről világ-szerte ismert márkává vált. Gladwell azt állítja, hogy a kulcs az úgynévezett *Összekötőkön* műlik. Egy összekötő tipikusan hatalmas társasági kapcsolatrendszerrel rendelkezik, mely kapcsolatok többnyire gyenge kötelék csupán, emellett több mikrovilágba és szubkultúrába bejáratos. Az ilyen összekötőket kell megnyernie egy terméknek és ennek szinte automatikus következménye, hogy elterjed a termék és divattá válik. Duncan Watts ausztrál fizikus-matematikus-szociológus megkérdőjelezte Gladwell állítását, szerinte ugyanis nem az a meghatározó, hogy egy divat az összekötőktől indul-e, hanem az, hogy a közösség (akár az egész társadalom, akár csak egy szűkebb baráti kör) készen áll-e az adott divat elterjedésére. Tehát nem a forrás számít és annak befolyása, hanem az, hogy általánosságban a közösség egésze akarja-e a terméket.

2.2.1. Trendterjedés, mint útvonalválasztási probléma

A trendterjedés nagyon hasonlít a vírusterjedésre, hiszen maga a két eredeti modell is szin- te egy az egyben megfeleltethető egymásnak. Természetesen a vírusterjedés esetén említett kiegészítéseket nem lehet egy az egyben átültetni és a trendterjedés terminológiájával le-

írni, de az alapmodell három csoportját (S , I , R) itt is azonosíthatjuk. Világos, hogy az optimalizálás lokalizáltsága erre a modellre is jellemző, hiszen nincs olyan divat, ami egy bizonyos emberre akar csak hatni, sokkal inkább mindenki. Így ennek a problémakörnek is jellemző policy-ja lehet az *Unió-fedés*.

Útvonalválasztási szempontból a lényeges különbség, hogy a trendterjedésnél a csomópontoknak van döntő szerepe, nem az éleknek, azaz itt nem közös érdek, hogy egy divat elterjedjen (vagy akar az, hogy ne terjedjen el). Bárki mondhatja azt, hogy neki kifejezetten nem tetszik az elterjedő divat, ezért pl. lebeszéli az ismerőseit róla, vagy csak egyszerűen nem veszi át, így megállítja a terjedést, legalábbis a közvetlen közelében. Nyilvánvaló, hogy - bár az elterjedés kulcsa az, hogy a társadalom készen áll-e - egy *Összekötő* nagyban hozzájárulhat a sikeres terjesztéshez, így preferálhatjuk ezen központi csomópontokat. Emellett az is szempont lehet, hogy nem a csomópontok befolyását vesszük alapul, hanem pusztán a mennyiségeket és egy adott útvonalválasztáskor minél hosszabb útra törekszünk. Egy másik alkalmazható heurisztika, hogy ha két csomópont közötti élen sok divat terjedt át, akkor ezt az élet érdemes felhasználni, illetve megfordítva, ha egy élen eddig kevés divat terjedt át, akkor valószínűleg nem olyan a két végpont kapcsolata, amin a divatok átterjedhetnek, így érdemes elkerülni az ilyen éleket. Ezek alapján a következő policy-kat definiálom:

- **Összekötő-keresés:** A nagy súlyú csomópontok felé irányító policy.
- **Korai-elfogadó-keresés:** A minden átvevő csomópontok felé irányító policy.

Összekötő-keresés

Ez a policy azon csomópontokat részesíti előnyben, amelyeket a legbefolyásosabbnak ítélik. A vírus- és trendterjedés hasonlóságából adódóan itt is lokális optimalizálásról beszélhetünk, így minden a szomszédot fogjuk választani, aminek a legnagyobb a fokszáma. Az *Összekötő*-keresés algebrája a $\mathcal{O}((1, d), 0, \max, \geq)$. Az élek súlya: $\forall u, v \in V : w(u, v) = \max(\deg(u), \deg(v))$, tehát az összekötött csomópontok fokszámai közül a nagyobb és nyilván a több jobb. Az átjáratatlanságot a 0 súly jelzi, míg több él súlyát a maximumuk határozza meg.

Korai-elfogadó-keresés

A Korai-elfogadó-keresés policy azon éleket részesíti előnyben, melyeken az eddigi tapasztalat alapján sok divat átterjedt már. Ennek lényege nyilván az, hogy egy ilyen él olyan személyek közötti kapcsolatot jelöl, akik pl. megbíznak a másik ítéletében, vagy az egyikük felnéz a másikra, így átvesz tőle viselkedési formákat. A policy algebrája a $\mathcal{K}(\mathbb{N}, -1, +, \geq)$. Az élek súlya azt jelzi, hogy eddig mennyi divat terjedt át rajta, így a több a jobb, és nyilván a -1 súlyú élen nem tudunk átmenni. Több él összeillesztését az élsúlyok összegével súlyozunk.

2.3. Útvonalválasztás az Interneten

A útvonalválasztás kutatói rendszerint informatikai-, vagy számításelméleti szakemberek, ezért a kutatási célok legnagyobb része számítógépes hálózatok útvonalválasztásának vizsgálata, javítása. Ezen a területen általában az Internet gerinchálózatát szokták vizsgálni, hiszen ott jelentkeznek azok a skálázhatósági- és menedzsmentproblémák, amik az Internet 1970-es évekbeli tervezéséből származnak és ami miatt azt mondhatjuk, hogy az Internet hibáinak toldozása-foldozása már nem elég és alapvetően új megoldások szükségesek.

Két alapvetően eltérő koncepció létezik a számítógépes hálózatok útvonalválasztásában, a tapasztalatok alapján előre becsült forgalmi viszonyoknak megfelelő centralizált kialakítás; illetve az aktuális forgalmi helyzet állandó figyelése alapján, az annak legjobban megfelelő irányítás. Legtöbbször az utóbbit valósítják meg, méghozzá elosztott változatban. Az Internet tervezése során több olyan szempontot is figyelembe vettek, melyeket az előző két alfejezetben nem tudunk tárgyalni. Ilyen például a routerekben lévő útvonaltáblák minél kisebb mérete (kisebb tár, olcsóbb és gyorsabban működő csomópont, illetve kisebb routingforgalom), emellett fontos szempont a routingforgalom minimalizálása, a robusz-tusság (hibás tábla esélyének minimalizálása: „fekete lyuk”, hurok, oszcilláció elkerülése) és végül az optimális útvonalak kijelölése (természetesen az út optimalitása az igénytől függ). Az egyik legfontosabb igény az elosztott működés, a decentralizáltság volt azért, hogy az útvonalválasztást a csomópontok végezzék.

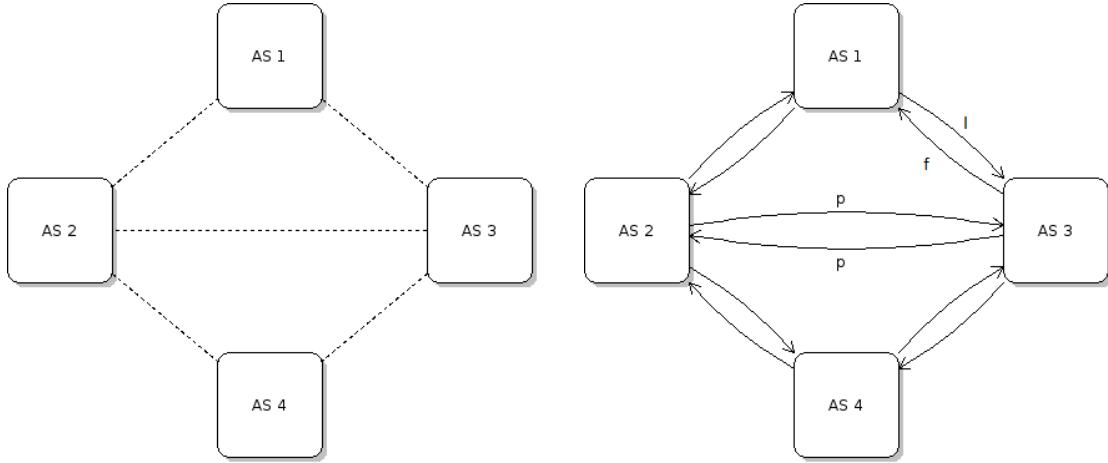
Ahogy fejlődött a világ, újabb feladatok és problémák jelentek meg, amelyet a szakemberek a hálózat tervezésekor még nem láttak előre. Olyan új technológiák jelentek meg, mint a mobiltelefonok és a hordozható számítógépek és velük együtt az igény a mobil routingra, azaz arra, hogy ne csak a számítógép, de az internet hozzáférés is legyen hordozható. Problémát jelent a multicast routing is, amikor egy adott csomagot több címettől szeretnénk eljuttatni. Ezen problémák hatékony megoldása kritikus feladat, alapvetően új megoldásokat kívánnak.

Emellett pusztán a hálózat mérete is feszegeti az Internet teljesítőképességének jelenlegi határait. Mára az Internet óriásira nőtt és a 1990-es évek vége óta a minden nap élet meghatározó részévé vált. 2000 és 2009 között 394 millióról 1,84 milliárd nőtt az Internet kapcsolattal rendelkezők száma⁴. Ma már több, mint 2,4 milliárd⁵ felhasználója van az Internetnek, és 2010-ben 12,5 milliárd Internetre csatlakoztatott eszköz volt. Ez azt jelenti, hogy minden élő emberre jutott 1,84 készülék. 2020-ra az Internetre kapcsolt eszközök száma elérheti az 50 milliárdot [21].

⁴Market Information and Statistics, International Telecommunications Union, <http://www.itu.int/ITU-D/ict/statistics/>

⁵2.405.518.376 - 2012. második negyedév, <http://www.internetworldstats.com/stats.htm> 2014. 05. 09

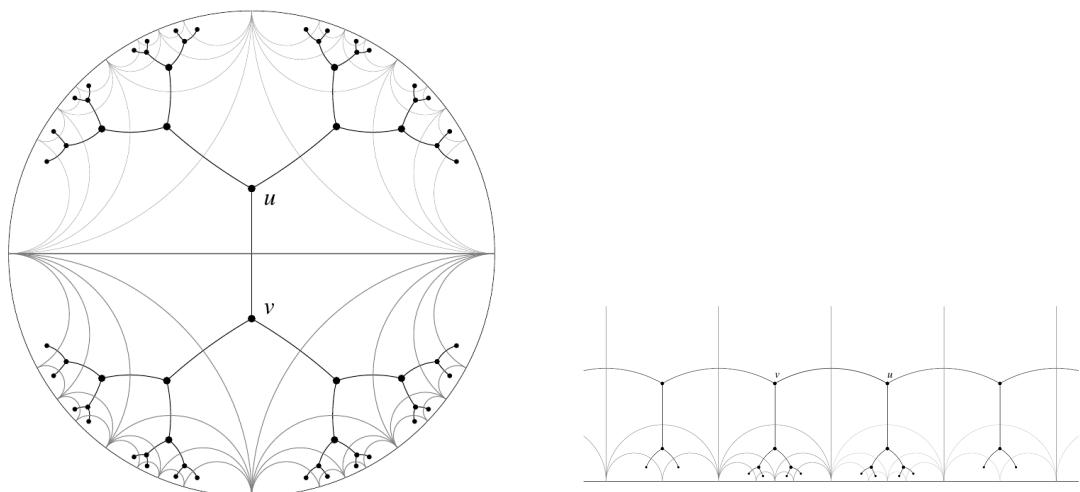
Útvonalválasztási szempontból az Internet routing egy klasszikus feladat, amikor a legszélesebb-, legrövidebb- és legmegbízhatóbb utakat keressük. Ezen policy-k algebráit már bemutattam, megtalálhatók a az 1.5.1 táblázatban.



2.1. ábra. A BGP egyszerűsített képe és a völgymentesség szerinti irányítás.

Érdemes megvizsgálni a BGP szabályrendszerét: a BGP routing policy-ja többszintű. A legalsó szint, a legalapvetőbb policy a völgymentesség. Ez azt jelenti, hogy az útvonalválasztásnál elsődleges szempont, hogy semelyik AS-nek ne kelljen fizetni olyan forgalomért, ami csak áthalad rajta. Ha a hierarchiában lefelé mutató éleket l -lel, a felfelé mutató éleket f -vel, a peering kapcsolatokat pedig p -vel jelöljük, akkor minden routing során kijelölt útvonal csak a következő alakban írható le: néhány (akár nulla) f él, aztán maximum egy p él, utána pedig néhány (akár nulla) l él.

Emellett, ha felhasználjuk a gráfbeágyazás és kompakt routing eredményeit, akkor egy új távolságfüggvénnyel és a csomópontok koordinátázásával megvalósítható az Internet tartományszintű gráfjában is egy elakadásmentes mohó útvonalválasztás [22].



2.2. ábra. A hiperbolikus sík Poincaré-féle diszk modellje és a felső félsík modellje [23].

A \mathbb{H} hiperbolikus síknak többféle szemléltető modellje is van, ám a két legelterjedtebb a felső félsík modell és a Poincaré diszk modell, amiket a 2.2 ábrán láthatunk is. A Poincaré modellben \mathbb{H} -t egy egységkörrel reprezentálják: $x^2 + y^2 < 1$, a következő metrikával:

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}. \quad (2.4)$$

A felső félsík modellben \mathbb{H} -t a $\{\langle x, y \rangle \mid y > 0\}$ ponthalmaz írja le, ahol a metrika:

$$ds^2 = \frac{dx^2 + dy^2}{y^2}. \quad (2.5)$$

Mindkét esetben a \mathbb{H} pontjait komplex számokként kezeljük: $(x, y) \in \mathbb{R}^2 : z = x + yi$. Ezek alapján két új policy-t mutatok be:

- Hiperbolikus-távolság: Az elakadásmentes mohó útvonalválasztás policy-ja.
- Völgumentesség: A BGP alapvető, elsődleges policy-ja.

Hiperbolikus-távolság

A hiperbolikus síkra ágyazott Internet gráf minden pontja egy (x, y) koordinátapárral leírható. A policy algebrája a $\mathcal{H}(\mathbb{R}^+, \infty, f_{\mathbb{H}}, \leq)$, ahol $f_{\mathbb{H}}$ egy viszonylag bonyolult függvény, definiálásához szükséges lenne több, itt nem részletezett ismeret a hiperbolikus geometria téma köréből (bővebben ld. [24] 2. fejezetét).

Völgumentesség

A völgumentesség policy-nak az algebrája a $\mathcal{V}(\{f, l, p\}, \phi, \oplus_{\mathcal{V}}, \preceq)$, ahol a $\oplus_{\mathcal{V}}$ a 2.3.1 táblázat szerinti⁶. Az előbbi szabály másik megfogalmazásban azt jelenti, hogy sem l -t, sem p -t nem követhet sem p , sem f .

2.3.1. táblázat. A $\oplus_{\mathcal{V}}$ [8].

\oplus	f	l	p
f	f	l	p
l	ϕ	l	ϕ
p	ϕ	p	ϕ

⁶ Egy adott súlyú (típusú) úthoz hozzá akarnánk venni egy élet, akkor az út minden súlyúvá (típusúvá) válna.

2.4. Egyéb algebrák

A minél részletesebb vizsgálat érdekében definiálok olyan policy-kat is, melyek vagy minden eddigi modellben használható lenne, vagy egyikben sem, így eddig nem volt lehetőségem bemutatni.

A fejezet legelején felvázolt két dimenzió (optimalizálás, érdekek) mellett a harmadik karakterizálási lehetőség az időbeli lefolyás. Ha egy útvonalválasztási probléma tárgyalása során figyelembe vesszük az időt is, mint befolyásoló tényezőt, egy olyan új dimenziót ragadunk meg, mely minden modell routing-ját képes befolyásolni: ez az *Időfüggés* policy. A vírus-terjedsnél elég csak arra gondolni, hogy télen könnyebben tud terjedni a fertőzés, és hasonlóan nyáron nagyobb valószínűséggel terjed egy fürdőruhádivat. Fontos azonban, hogy ez csak egy mellékes faktor, azaz egy alap policy mellett van értelme figyelembe venni az időpontot is. Az Internetes útvonalválasztás során időfüggést tapasztalhatunk, ha pl. egy terheléselosztó rendszer egy adott kérést másodpercenként váltakozva egyszer kiszolgál, egyszer pedig újrapróbálkozásra szólít fel.

Másik érdekes policy, a trendterjedsnél említett leghosszabb út policy. Ennek olyan esetben van értelme, amikor nem a célba érkezés a legfontosabb, hanem maga az út. A vírus- és trendterjedsnél ez azért volt lényeges, hogy minél több embert elérjünk, de pl. a hangyák is így építik ki a bolyban az utakat, hogy egy esetleges betolakodó minél nagyobb valószínűséggel eltévedjen.

Leghosszabb-út

A leghosszabb-út policy adja az elérhető leghosszabb utat. Ebben az esetben az \mathcal{L} algebra az $(1, 0, +, \geq)$ négyes. Ebben a policy-ban minden él konstans 1 súlyú, egy út súlya éppen az élszáma.

Időfüggés

Az Időfüggés policy lényege, hogy az időpontot⁷ figyelembe véve néha átjárhatatlan egy-egy él. Ehhez szükséges egy alap policy is, amelyet ez ki tud egészíteni: \mathcal{A} . Jelölje T_{e_i} az időpontok egy olyan halmozát, melyben \mathcal{I} nem enged át forgalmat az e_i élen. (A T_{e_i} lehet akár egy $[t_0, \infty)$ intervallum is.) Így az \mathcal{I} algebra: $(W_{\mathcal{A}}, \phi_{\mathcal{A}}, f, \preceq_{\mathcal{A}})$, ahol

$$f(e_1, e_2, t) = \begin{cases} \phi_{\mathcal{A}} & \text{ha } t \in T_{e_1} \cup T_{e_2} \\ e_1 \oplus_{\mathcal{A}} e_2 & \text{különben} \end{cases}$$

Azaz, alkalmas időpontban teljes egészében az \mathcal{A} policy érvényesül, azonban minden élen bizonysos időpontokban nem lehet átmenni.

⁷A lépték mértéke problémafüggő: lehet naponkénti éves periódusokkal, lehet

2.5. Összefoglaló

Ebben a fejezetben megvizsgáltam a hálózatkutatás szempontjából alapvető modelleket, melyeket a lokális- vagy globális optimalizálás és a közös- vagy egyéni érdekek követése tulajdonságok alapján karakterizáltam. Bemutattam a fertőző betegségek vizsgálatára használt matematikai modellt, megvizsgáltam a téma kör útvonalválasztási kérdéseit és ki-jelöltet a két, a problémakört jól jellemző policy-t: a Fertőzési-határ és az Unió-fedés policy-ket, illetve ezek algebráit: $\mathcal{F} = ((0, 1], 0, \max, \geq)$ és $\mathcal{U} = (\mathbb{N}, \infty, f, \leq)$.

Rávilágított a trend- és a vírusterjedés hasonlóságaira és különbségeire útvonalválasztási szempontból és definiáltam két policy-t, a Összekötő-keresés-t és a Korai-elfogadó-keresés-t. Ezen policy-k algebrái: $\mathcal{O} = ((1, d), 0, \max, \geq)$ és $\mathcal{K} = (\mathbb{N}, -1, +, \geq)$.

Végül megvizsgáltam a már ismertetett policy-kon (ld. az 1.5. alfejezet) kívül az Internet tartományszintű gráfjának az alapszabályát, a Völgymentességet, illetve felvázoltam a hiperbolikus térbe ágyazás - általa pedig az elakadásmentes mohó útvonalválasztás - lehetőségét: $\mathcal{V} = (\{f, l, p\}, \phi, \bigoplus_{\mathcal{V}}, \preceq)$ és $\mathcal{H} = (\mathbb{R}^+, \infty, f_{\mathbb{H}}, \leq)$.

3. fejezet

A modellezési keretrendszer

3.1. Specifikáció

3.1.1. A bemenet

Az adatok előfeldolgozása, tisztítása

A domain szűkítése

3.1.2. A szimuláció

Utak generálása

3.1.3. A kimenet

3.2. A szimulátor

3.2.1. A matematikai struktúrák implementálása

3.2.2. A felhasznált algoritmusok elemzése

3.3. A szimulációs eredmények kiértékelése

3.3.1. Az eredmények validálása

3.3.2. Gráfok összehasonlítása

3.3.3. Az adatok megjelenítése

3.4. Valós hálózatok vizsgálata a keretrendszer segítségével

A 2. fejezetben definiált algebrák modellezése után valós hálózatokon vizsgálhatjuk a viselkedésüket, hogy hogyan alakítanak ki utakat. Ezután tetszőleges algebrák lexikografikus szorzatát is szimulálhatjuk, így még pontosabb és változatosabb modellt kapunk.

A valós hálózatokon kialakult útvonalak és a policy primitívek által meghatározott útvonalak összehasonlítására több megoldás is létezik, az egyik ilyen, hogy minden forrás-cél párra megvizsgálom, hogy a valós és a szimulált út mennyire hasonlít a lépésszám és az érintett köztes csomópontok alapján, illetve az adott p. primitív út nevezhető-e jobbnak az

adott problémakörben, esetleg egy alternatív útnak, vagy akár rosszabb-e, mint az eredeti útvonal.

Emellett a hálózatok egészét is összehasonlítom statisztikai módszerekkel: fokszámeloszlás, élek száma, összefüggőség. Ezzel feltárhatóak olyan változások, melyek a pont-pont kapcsolatok vizsgálata során nem vettünk észre, pl. lehetséges, hogy a valóságban skálafüggetlen hálózatnak a fokszámeloszlása a szimuláció során már sokkal kiegyensúlyozottabb lesz. Az Internet tervezésekor fontos szempont volt a robusztusság, amit a gráf pont- ill. él-összefüggésével tudunk mérni, nyilván egy olyan hálózat, amiben a routing táblák mérete kicsi, de nagy a szétszakadás veszélye, nem jobb, mint az eredeti, valós hálózat, sőt ez egy indok is lehet, hogy miért éppen azokat a szabályokat használták az AS-k, amiket titkolnak a nyilvánosság elől.

3.5. Összefoglaló

4. fejezet

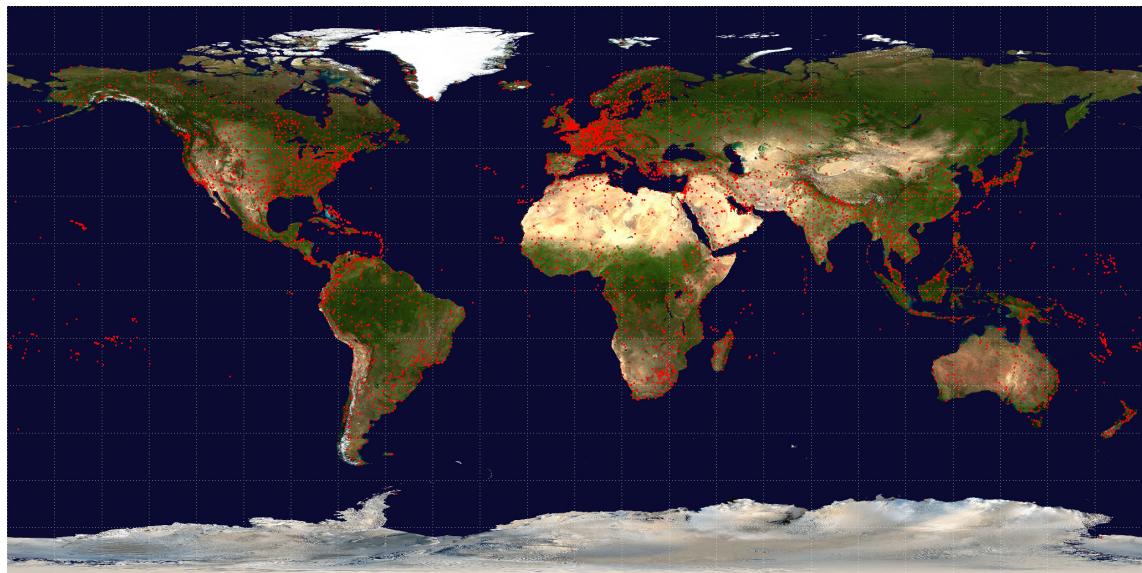
Repülési hálózat vizsgálata

Az openflights.org oldalon elérhető repülési adatbázis feldolgozása és szimulációs elemzése az egyik olyan feladat, amit a következő félév végéig fogok elvégezni. Az oldal lehetőséget biztosít interaktív keresésre, regisztráció után pedig statisztikákat készíthetünk saját utazásainkról. Emellett elérhető a teljes adatbázisa is, amit többek között hivatalos forrásokból, illetve a felhasználóktól gyűjt.

4.1. Az adatok

Az openflights-on elérhető adatbázis világszerte mindenhonnan gyűjt és tárol reptéri, repülési társasági és repülési út adatokat. Közel 8000 reptér, 6000 társaság és 60000 repülési út érhető el. Részletes adatok a 4.1.1, a 4.1.3, és a 4.1.2 táblázatokban.

4.1.1. Repterek



4.1. ábra. Az OpenFlights adatbázisában szereplő repterek.

A rendszer tárolja a repterek *nevét*; a *várost*, ahol található; a *szélességi-* és *hosszúsági fokokat*, illetve a *tengerszint feletti magasságot*. Emellett elérhető a repülési hatóságok által

használt 3 karakter hosszú *IATA*¹/*FAA*² azonosító is és az OpenFlights által használt decimális azonosító is. A legutolsó, 2014 januári frissítéskor az OpenFlights adatbázisa 7733 repteret tartalmazott világszerte, ezt mutatja a 4.1 ábra.

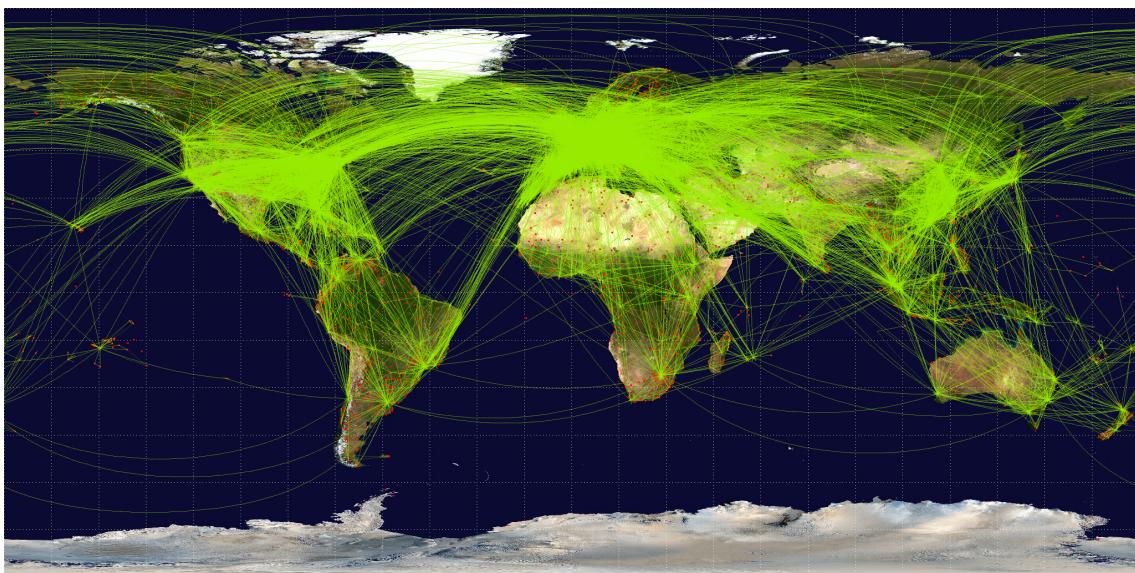
4.1.1. táblázat. Az OpenFlights adatbázisában elérhető reptéri adatok.

Attribútum	Leírás
Airport ID	Egyedi OpenFlights azonosító.
Name	A reptér neve.
City	Az a város, amit a reptér „kiszolgál”.
Country	Az ország vagy terület neve, ahol reptér található.
IATA/FAA	3 betűs FAA kód az USA-beli reptereknek vagy 3 betűs IATA kód, minden más esetben.
ICAO ³	4 betűs ICAO kód.
Latitude	A reptér szélességi foka: decimális szám (fokban mérve), általában 6 szignifikáns jegyig. A Déli féltekén negatív, az Északon pozitív.
Longitude	A reptér hosszúsági foka: decimális szám (fokban mérve), általában 6 szignifikáns jegyig. A Nyugati féltekén negatív, a Keleten pozitív.
Altitude	A tengerszint feletti magasság, lábban mérve (1 láb ~ 0.30 méter).

¹International Air Transport Association: Nemzetközi Légi Szállítási Szövetség

²Federal Aviation Administration: Szövetségi Légügyi Hivatal (USA)

4.1.2. Repülési útvonalak



4.2. ábra. Az OpenFlights adatbázisában szereplő repülési útvonalak.

A repülési útvonalak leírása öt attribútumot tartalmaz: az *üzemeltető repülési társaság* (név és OpenFlights azonosító); a *forrás reptér*; a *cél reptér*. Emellett feljegyzik a *megállások számát*⁴, illetve az ebben a viszonylatban használt *repülőgép típusokat*. Ha két reptér között oda-vissza is van járat, akkor ez az adathalmazban két bejegyzésként jelenik meg. Jelenleg az OpenFlights adatbázisa 59637 útvonalat tartalmazott 3285 reptér és 531 társaság⁵ között világszerte, ahogy a 4.2 ábra is mutatja.

4.1.2. táblázat. Az OpenFlights adatbázisában elérhető adatok a repülési útvonalakról.

Attribútum	Leírás
Airline	Az üzemeltető társaság 2 betűs (IATA) vagy 3 betűs (ICAO) kódja.
Airline ID	Az üzemeltető társaság egyedi OpenFlights azonosítója.
Source airport	A forrás reptér 3 betűs (IATA) vagy 4 betűs (ICAO) kódja.
Source airport ID	A forrás reptér egyedi OpenFlights azonosítója.
Destination airport	A cél reptér 3 betűs (IATA) vagy 4 betűs (ICAO) kódja.
Destination airport ID	A cél reptér egyedi OpenFlights azonosítója.
Codeshare	Igaz, ha a járat „codeshare”, azaz nem utasszállító járat, különben üres.
Stops	A megállások szám (nem az átszállásokat jelenti, a „0” jelenti a direkt járatot).
Equipment	A viszonylatban használt repülőgép típusok 3 betűs kódjai.

⁴Az összesen 59637 bejegyzésből 6 esetben van 1 megállás és 1 esetben 2 megállás.

⁵Ugyan a légitársaság és reptér adatbázisokban ennél jelentősen több bejegyzés van, ám azok között vannak nem aktív társaságok, illetve bezárt repterek.

4.1.3. Repülési társaságok

A repülési társaságokról tárolt adatok többek között tartalmazzák a hivatalos *IATA azonosítót*; a cég *nevét*; az *országot*, ahol be van jegyezve; illetve, hogy *aktív-e* még a társaság. Emellett természetesen itt is definiáltak egy saját (OpenFlights) *azonosítót*.

4.1.3. táblázat. Az OpenFlights adatbázisában elérhető adatok a repülési társaságokról.

Attribútum	Leírás
Airport ID	Egyedi OpenFlights azonosító.
Name	A társaság neve.
Alias	A társaság egyéb megnevezése.
IATA	2 betűs IATA kód.
ICAO	3 betűs ICAO kód.
Callsign	A társaság hívójele.
Country	Az ország vagy terület neve, ahol a társaság be van jegyezve.
Active	Igaz, ha a társaság jelenleg is, vagy nemrég még működött (nem megbízható).

4.2. A szimuláció

4.2.1. Az adatok előfeldolgozása

A modellezendő hálózat első megközelítésben az összes csomópontból alkotott K_{7733} teljes gráf. Ez természetesen feleslegesen nagy vizsgálandó hálózatot jelentene ($\binom{7733}{2} = 29.895.778$ élű gráf.), emellett az olyan, leíró jellegű jellemzők, mint a valódi hálózat átmérője, vagy pont- és él-összefüggőségei elvésznek. Lévény a repülési hálózat skálafüggetlen - sok olyan csomópont van, ahova csak adott, kis számú, közeli csomópontokból indulnak járatok, így nem érdemes összekötni távoli csomópontokkal. Ezenkívül a repülésnél vannak olyan szempontok, amelyeket az útvonalválasztásnál figyelembe vesznek, ám én ezek pontos információk hiányában nem tudom kezelni - pl. csak adott hosszúságú utat tehet meg egy adott típusú repülőgép az üzemanyagtartálya méretétől függően.

4.2.2. A szimuláció megtervezése

A fentiek miatt a szimulált hálózatot a következőképpen határozzom meg: pontjai a reperek, élei pedig, az adatbázisban szereplő össze, megállás nélküli repülőút. Így magától értetődően csak olyan élek lesznek a szimulációban, amelyek „átrepülhetők”, hiszen volt már, hogy átrepülték. Emellett pedig ez a hálózat tükrözi a valóságot, hiszen feltehetjük, hogy csak azon élek nincsenek behúzva, amelyeket repüléstechnikai okokból nem is lehet, különben valaki, valamikor már repült volna azon az úton.

Az útvonalválasztást a 10 legjelentősebb légitársaság alapján szimulálom, mert a 10 legtöbb járatot üzemeltető társaság az utak 24%-át adják. A policy vezérelt útvonalválasztást az összes olyan pontpár között lefuttatom, amely ezen 10 társaság célpontja.

4.2.3. A vizsgálandó metrikák, az élek súlyozása

A fejezet elején említett két megközelítést fogom alkalmazni, azaz a pont-pont kapcsolatokat is összehasonlítom és a globális hálózat mutatóit is. Az élek súlya annak megfelelően változik, hogy milyen policy algebráját szimulálom. Emellett a repterek GPS koordinátái alapján meghatározom az összekötött repterek fizikai távolságát. Mivel a légtérben nincsen torlódás az út szélességének definiálásának nincs értelme.

A globális hálózati metrikák közül a fokszámeloszlást fogom vizsgálni, illetve a hálózat átmérőjét és a pont- és él-összefüggőségét. A fokszámeloszlás és az átmérő összehasonlításra azért van szükség, mert ezek alapvető karakterisztikáját adják a hálózatnak. A pont- és él-összefüggőséget pedig azért, mert ezek lehetnek olyan tényezők, amik a repülőtársaságok útvonalválasztása mögött meghúzódhatnak.

4.3. Az eredmények

4.4. Összefoglaló

5. fejezet

Összefoglalás

A munka első felét elvégeztem, áttekintettem a hálózatkutatás eredményeit, különös tekintettel az útvonalválasztás matematikai kérdésére.

Az 1. fejezetben áttekintettem a szakirodalmat, összeszedve a legfontosabb állomásokat. Rámutattam, hogy az Internet AS-szintű topológiáján kívül, a más tudományterületekről származó problémák útvonalválasztásáról sem tudunk sok minden és ezért szükséges egy olyan eszköz, ami a policy-feltárást feladatát - általános esetben is - hatékonyan el tudja látni. Ehhez definiáltam a *routing algebrát*, bemutattam a legfontosabb tulajdonságait és műveleteit, emellett a legszélesebb körben használt policy-k algebráit is ismertettem.

A 2. fejezetben megvizsgáltam a hálózatkutatás szempontjából alapvető modelleket, melyeket a lokális- vagy globális optimalizálás és a közös- vagy egyéni érdekek követése tulajdonságok alapján karakterizáltam. Bemutattam a fertőző betegségek vizsgálatára használt matematikai modellt, megvizsgáltam a téma kör útvonalválasztási kérdéseit és kijelöltettem a két, a problémakört jól jellemző policy-t. Emellett rávilágított a trend- és a vírus terjedési útvonalválasztási szempontbeli hasonlóságaira és különbségeire és definiáltam két új trendterjedési policy-t.

Megvizsgáltam a már ismertetett policy-kon (ld. az 1.5. alfejezet) kívül az Internet tartománszintű gráfjának az alapszabályát, a völgymentességet, illetve felvázoltam a hiperbolikus térbe ágyazás - általa pedig az elakadásmentes mohó útvonalválasztás - lehetőségét.

A 4. fejezetben bemutattam a vizsgálandó valós hálózatot. Az adatok feldolgozása után meghatároztam a vizsgálandó hálózatot. Definiáltam a vizsgálandó metrikákat, amik mentén össze tudom hasonlítani a valós- és a policy vezérelt útvonalválasztás által meghatározott útvonalakat és hálózatokat. minden pontpárra megvizsgálom az eredeti és a szimulált út különbségeit (távolság, lépésszám). A valós és a szimulált hálózatok statisztikai összehasonlítását is elvégzem, melyben figyelmet fordítok a fokszámeloszlás, a hálózatok átmérőjének és a pont- ill. él-összefüggőségeinek összehasonlítására.

Ábrák jegyzéke

2.1.	A BGP egyszerűsített képe és a völgymentesség szerinti irányítás.	24
2.2.	A hiperbolikus sík Poincaré-féle diszk modellje és a felső félsík modellje [23].	24
4.1.	Az OpenFights adatbázisában szereplő repterek.	30
4.2.	Az OpenFights adatbázisában szereplő repülési útvonalak.	32

Irodalomjegyzék

- [1] Duncan J. Watts, *Networks, Dynamics, and the Small-World Phenomenon*. American Journal of Sociology, Volume 105, Issue 2, 493-527. oldal, 1999.
- [2] Döbrei G., *Nagyméretű hálózatok evolúciója - Trendterjedés vizsgálata szimulációs eszközökkel*. Beszámoló, BME TMIT, Hálózatok és szolgáltatások, Önálló laboratórium 1., 2013.
- [3] Ramesh Govindan, Hongsuda Tangmunarunkit, *Heuristics for Internet Map Discovery*. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE Volume 3, 1371-1380. oldal, 2000.
- [4] Hyunseok Chang, Sugih Jamin, Walter Willinger, *Inferring AS-level Internet Topology from Router-Level Path Traces*. In Proceedings of SPIE ITCom 2001, 2001.
- [5] Feng Wang, *On Inferring and Characterizing Internet Routing Policies*. Proceeding IMC '03 Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement, 15-26. oldal, 2003.
- [6] Hongsuda Tangmunarunkit, Ramesh Govindan, Scott Shenker, Deborah Estrin, *The Impact of Routing Policy on Internet Paths*. INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE Volume 2, 736-742. oldal, 2001.
- [7] J. L. Sobrinho, *Algebra and Algorithms for QoS Path Computation and hop-by-hop Routing in the Internet*. IEEE/ACM Transactions on Networking (TON), Volume 10., Issue 4, 541-550. oldal, 2002.
- [8] Rétvári G., Gulyás A., Heszberger Z., Csernai M., Biró J., *Compact Policy Routing*. Proceedings of the ACM Principles of Distributed Computing, 2011.
- [9] J. L. Sobrinho, *Network routing with path vector protocols: theory and applications*. In SIGCOMM '03, 49-60. oldal, 2003.
- [10] T. Griffin, J. L. Sobrinho, *Metarouting*. SIGCOMM '05, 1-12. oldal, 2005.
- [11] A. Gurney, T. Griffin, *Lexicographic products in metarouting*. In Network Protocols, IEEE International Conference on, 113-122. oldal, 2007.

- [12] Zheng Wang, Jon Crowcroft, *Quality-of-service routing for supporting multimedia applications*. IEEE Journal of Selected Areas in Communications, Volume 14, Issue 7, 1228-1234. oldal, 1996.
- [13] G. Apostolopoulos, R. Guerin, S. Kamat, S. K. Tripathi, *Quality of service based routing: A performance perspective*. In SIGCOMM, 17-28. oldal, 1998.
- [14] Qingming Ma, P. Steenkiste, *On path selection for traffic with bandwidth guarantees*. In Proceedings of the 1997 International Conference on Network Protocols (ICNP '97), 191. oldal, 1997.
- [15] P. Bose, P. Morin, I. Stojmenovic és J. Urrutia, *Routing with Guaranteed Delivery in Ad-hoc Wireless Networks*. Wireless Networks, Volume 7, Issue 6, 609-616. oldal, 2001.
- [16] D. J. Daley, J. Gani, *Epidemic Modelling: An Introduction*. Cambridge University Press, 2005.
- [17] D. Bernoulli and S. Blower, *An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it*. Reviews in Medical Virology, Volume 14, Issue 5, 275-288. oldal, 2004.
- [18] W. O. Kermack, A. G. McKendrick, *Contributions to the mathematical theory of epidemics*. Bulletin of Mathematical Biology, Volume 53, Issue 1, 33-55. oldal, 1927.
- [19] Csaba Z. L., Pál J., *Az utálat és az iskolai agresszió empirikus hálózatelemzése*. VII. HUNNET konferencia, Budapesti Corvinus Egyetem, 2010.
- [20] M. Gladwell, *Fordulópont: Ahol a kis különbségekből nagy változás lesz*. HVG Kiadói Rt., 2007, ISBN: 9789639686007.
- [21] D. Evans, *The Internet of Things - How the Next Evolution of the Internet Is Changing Everything*. Cisco Internet Business Solutions Group (IBSG), 2011.
- [22] Döbrei G., *Gráf-beágyazási technikák kommunikációs hálózatokban*. BSc szakdolgozat, Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformációk Tanszék, 2012.
- [23] R. Kleinberg, *Geographic Routing Using Hyperbolic Space*. INFOCOM 2007. 26th IE-EE International Conference on Computer Communications. IEEE, 1902-1909. oldal, 2007.
- [24] W. P. Thurston, *Three-Dimensional Geometry and Topology*. Princeton University Press, 1997.

Függelék

F.1. Számításelméleti alapok

F.1.1. Definíció (O , Ω , Θ). $f, g : \mathbb{N} \rightarrow \mathbb{R}$

- Ordó: $f(n) = O(g(n))$, ha $\exists c > 0, n_0 > 0 : |f(n)| \leq c|g(n)|, \forall n > n_0$.
- Omega: $f(n) = \Omega(g(n))$, ha $\exists c > 0, n_0 > 0 : |f(n)| \geq c|g(n)|, \forall n > n_0$
- Teta: $f(n) = \Theta(g(n))$, ha $f(n) = O(g(n))$ és $f(n) = \Omega(g(n))$, azaz $\exists c_1, c_2 > 0, n_0 > 0 : c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|, \forall n > n_0$

F.1.1. Megjegyzés. A nagy ordós jelölésből szokás kihagyni a $\log n$ -es szorzót, mint-hogy egy decimális számot mindenkorban át kell konvertálni $\log n$ bitre, így egy algoritmus lépésszámánál vagy (lokális/globális) memóriaigénynél biztosan nem tudjuk megspórolni: $\tilde{O}(n) = O(n \log n)$.

F.1.2. Definíció (k-szorosan pont-összefüggő gráf 1.). Egy G gráf k -szorosan pont-összefüggő, ha tetszőleges k -nál kisebb elemszámú csúcshalmazát elhagyva összefüggő gráfot kapunk, és $|V(G)| \geq k$.

F.1.3. Definíció (k-szorosan él-összefüggő gráf 1.). Egy G gráf k -szorosan él-összefüggő, ha tetszőleges k -nál kisebb elemszámú élhalmazát elhagyva összefüggő gráfot kapunk.

F.1.4. Definíció (összefüggő gráf átmérője). A G összefüggő gráf átmérője a két legtávolabbi csúcsának távolsága, azaz az összes csúcspár közötti legrövidebb utak közül a leghosszabbnak a hossza. Jelölje $P(u, v)$ az u és v csúcsok közötti utak halmazát, és $l(p)$ a út hosszát, ekkor a gráf átmérője:

$$D_G = \max_{u, v \in V} \min_{p \in P(u, v)} l(p)$$

F.2. Az útvonalválasztás teljes modellje

Ahhoz, hogy felépíthessük a routing teljes modelljét, a routing algebrákon (az 1.2. rész) kívül egy routing függvényre van szükség. Ebben a modellben a csomagok (ahogyan a valóságban is) hasznos teherből (payload) és egy header-ből állnak. Ha adott az \mathcal{A} routing algebra és a G gráf, akkor a policy routing függvény az $R : \mathbb{N} \times \mathbb{N} \mapsto \mathbb{N} \times \mathbb{N}$ leképezés, a csúcsok $L_V : V \mapsto \mathbb{N}$ címkézésével és az élek $L_E : E \mapsto \mathbb{N}$ címkézésével, a következőképpen: minden s, t pontpárra egymás után alkalmazva R-t:

$$(h_{i+1}, l_{i+1}) = R(v_i, h_i), \forall i = 1, \dots, k-1$$

megadja a preferált $p_{st}^* = (s =)v_1, v_2, \dots, v_k (=t)$ utat \mathcal{A} szerint a megfelelő $l_{i+1} = (v_i, v_{i+1})$ él-címkékkal, ahol h_1 egy alkalmas kezdő header. Azt mondjuk, hogy R megvalósítja az \mathcal{A} policy-t G -n. Még néhány óvintézkedést meg kell tennünk, hogy a címkékkel nehogy több routing információt kelljen kódolni a szükségesnél, azaz *clogn*¹, valamilyen alkalmas c konstanssal.

Ezek alapján tehát a routing a következő képpen történik: Ha egy u csomópont kap egy üzenetet h header-rel, akkor egyszerűen kiszámolja az $R_u(h)$ lokális routing függvény értékét: $R_u(h) = R(u, h)$, hogy megkapja az új header-t, h' -t és a kimenő portot, l -t. Ezután u beállítja a csomag header-jének h' -t és továbbküldi l -en keresztül. A routing függvénytel már könnyen megadható egy hálózat minden csomópontjának a lokális memóriaigénye ahoz, hogy egy adott routing policy-t valósítson meg:

F.2.1. Definíció (Routing policy megvalósításához szükséges lokális memória).

Az \mathcal{A} routing policy megvalósításához szükséges $\mathcal{M}_{\mathcal{A}}$ lokális memória:

$$\mathcal{M}_{\mathcal{A}} = \max_{G \in \mathcal{G}_n} \min_{R \in \mathcal{R}} \max_{u \in V} \mathcal{M}_{\mathcal{A}}(R, u),$$

ahol $\mathcal{M}_{\mathcal{A}}(R, u)$ az R_u lokális routing függvény kódolásához szükséges bitek minimális száma, \mathcal{R} azon policy routing függvények halmaza, melyek megvalósítják \mathcal{A} policy-t valamely G gráfon, és \mathcal{G}_n az összes n csúcsú gráfok halmaza.

F.3. Tételek a routing algebrák témakörében

F.3.1. Tétel. *Ha \mathcal{A} algebra kiválasztó (S) és monoton (M), akkor tömöríthető.*

F.3.2. Tétel. *Ha \mathcal{A} algebra szigorúan monoton (SM), akkor nem tömöríthető.*

Ennél egy általánosabb téTEL, aminek következménye az előző téTEL:

F.3.3. Tétel. *Ha \mathcal{A} algebra tartalmaz egy delimitált (D), szigorúan monoton (SM) subalgebrát, akkor \mathcal{A} nem tömöríthető.*

¹ Ez a címek kódolásához szükséges információ.

F.4. A szimulátor technikai részletei

F.4.1. A választott platform

F.4.2. Fejlesztői megjegyzések