

Exaeskalarako sare-interkonexioen diseinurako helburu-aniztasuneko optimizazioa

Designing an Exascale Interconnect using Multi-objective Optimization

Jose A. Pascual*, Javier Navaridas

APT ikerketa-taldeko ikertzaileak, University of Manchester, UK

LABURPENA: Exaeskala errendimendua milioika kalkulu-nukleo osaturiko sistemak erabiliz lortuko da. Elementu horiek konektatzeko moduak (sare-topologia) izugarriko eragina du hainbat errendimendu ezaugarritan. Lan honetan, sare-topologiak diseinatzeko helburu-aniztasuneko optimizazioan oinarritutako ingurune bat proposatzen dugu, EBk finantzatuta ExaNeSt proiektuan garatzen ari garena. ExaNeSt sistemaren sarearen modulartasunari esker sare-topologia ezberdinak diseinatu ditzakegu hainbat errendimendu-helburu optimizatuz; esaterako, inguruko komunikazioak, hutségite-tolerantzia eta energia-kontsumoa. Topologiak sortzeko prozesua optimizazio bidez gauzatzen da (sare-topologiaren hainbat ezaugarri minimizatuz) teknika ebolutiboak erabilia. Simulazio bidezko emaitzen analisiak sortutako topologiek errendimendu-helburuak betetzen dituztela erakusten du. Gainera, sare-topologia ezagun batekin egindako konparazioan ikus daiteke gure proposamenak sortzen dituen sareek propietate topologiko hobek daukatela eta, aldi berean, errendimendu handiagoa lortzen dutela.

HITZ GAKOAK: Exaeskala errendimendua, helburu-aniztasun optimizazioa, teknika ebolutiboak, simulazioa.

ABSTRACT: Exascale performance will be delivered by systems composed of millions of interconnected computing cores. The way these computing elements are connected with each other (network topology) has a strong impact on many performance characteristics. In this work we propose a multi-objective optimization-based framework to explore possible network topologies to be implemented in the EU-funded ExaNeSt project. The modular design of this system's interconnect provides great flexibility to design topologies optimized for specific performance targets such as communications locality, fault tolerance or energy-consumption. The generation procedure of the topologies is formulated as a three-objective optimization problem (minimizing some topological characteristics) where solutions are searched using evolutionary techniques. The analysis of the results, carried out using simulation, shows that the topologies meet the required performance objectives. In addition, a comparison with a well-known topology reveals that the generated solutions can provide better topological characteristics and also higher throughput in almost all evaluated scenarios.

KEYWORDS: Exascale performance, multi-objective optimization, evolutionary techniques, simulation.

* **Harremanetan jartzeko / Corresponding author:** Jose A. Pascual. University of Manchester, Oxford RD, Manchester M13, UK. – joseantonio.pascual@ehu.eus – <https://orcid.org/0000-0001-5355-6537>.

Nola aipatu / How to cite: Pascual, Jose A.; Navaridas, Javier (2019). «Exaeskalarako sare-interkonexioen diseinurako helburu-aniztasuneko optimizazioa»; *Ekaia*, 35, 2019, 309-324. (<https://doi.org/10.1387/ekaia.19724>).

Jasoa: 15 maiatza, 2018; Onartua: 24 uztaila, 2018.

ISSN 0214-9001 - eISSN 2444-3255 / © 2019 UPV/EHU



Obra hau Creative Commons Atribución 4.0 Internacional-en lizentziapean dago

1. SARRERA

Super-konputazioko komunitatearen hurrengo erronka exa-flopak konputatzeko gai diren sistemak, exaeskala-sistemak, garatzea da. Horrelako errendimendua lortzeko aplikazio paralelo erraldoiak gauzatu behar dira, milioika nodoz (kalkulu-nukleoz) osaturiko sistemak erabiltzen dituztenak. Nodo horiek konektatzeko moduak, hau da, sare-topologiak, izugarriko eragina du sistemaren errendimendu teorikoan zein praktikoan. Hainbat sare-topologia erabili daiteke sistema paraleloak eraikitzeke; esaterako, toru-eta zuhaitz-topologiak edo berriki sortutako Dragonfly topologia [11]. Topologia horiek gaur egungo super-konputagailu handienetan erabiltzen dira; adibidez, *Blue Gene* familiakoetan [4], *Summit* super-konputagailua izango den horretan [13] eta *Cray* ko hainbat sistematan [6]. Azkenaldian, ausazko topologiak (irregularrak) proposatu dira, zeinek, agidanez, klasi-koek baino propietate hobekak dituzten [15].

European Exascale System Interconnect and Storage (ExaNeSt) [8, 9] proiektuan exa-eskalarako gai den sistema bat diseinatzen eta eraikitzen ari gara. Gure helburua kontsumo txikiko milioika ARM nukleo [2] erabil ditzakeen arkitektura bat garatzea da, tamaina handiko eta datu handiko problemak ebazte ditzakeena. Horren tamaina handiko sistemari eusteko interkonexio-sare berezi bat diseinatu behar dugu, errendimendu- eta erre-silientzia-maila beteta eta kostu-mugen barruan.

ExaNeSt-eko interkonexio-sareak hainbat maila dauzka bi multzotan sailkatzen direnak. Alde batetik, beheko mailak finkoak dira *back-plane* ak erabiltzen dituztelako. Bestetik, goiko mailak guztiz aldakorrek dira FPGA-k erabiltzen dituzten bideratzaileak [5] darabiltzatelako. Hori dela eta, konputazio- eta sare-elementuak era batean baino gehiagotan lotu daitezke ausazko topologiak osatzeko [14]. Malgutasun horri esker bai topologia zuzenak, bideratzaileen arteko zuzeneko konexioak erabiltzen dituztenak, bai zeharkako topologiak, bideratzaile osagarriak erabiltzen dituztenak, eraiki ditzakegu. Aldi berean, lehentasuna eman diezaiekegu ezaugarri zehatzak dauzkaten topologiei, esaterako, kostu-errendimendu edo energia-kontsumo maila zehatzak lortzen dituztenei edo hutsegite-tolerantzia baldintza zehatzak betetzen dituztenei. Hala ere, malgutasun horrek, sistemaren tamainarekin batera, soluzioen espazioa erraldoia izatea eragiten du, eta horregatik topologiak eskuz diseinatzea ezinezkoa da.

Artikulu honetan ezintasun horri aurre egiteko garatu dugun helburu-aniztasuneko optimizazio-ingurunea deskribatzen da, ezaugarri zehatzak dauzkaten topologiak diseinatzen lagunduko diguna. Ingurunea, bereziki, bi eboluzio-teknika ezagunetan oinarrituta dago eta problema honetarako gurutzatze- eta mutazio-eragile espezializatuak erabiltzen ditu. Nahiz eta ingurunea orokorra izan, lan honetan ExaNeSt-ek behar dituen ezaugarri jarriko diegu arreta, errendimendua eta hutsegite-tolerantzia maximi-

1. irudian ExaNeSt proiektuan garatutako arkitekturaren deskribapen orokorra erakusten da. ExaNeSt-en, kontsumo txikiko ARM nukleoak eta hardware-azeleragailuak erabiltzen dira konputatzeko, konputazio-elementuak deritzenak (KE), eta horietan oinarrituta eraikitzen da sistema osoa, mailaka antolatuta. Sei KE *backplane* batera konektatuta eraikitzen da lehenengo maila, hau da, txasisa (TX). *Backplane* batek banda-zabaleria handia ematen du (KE bakoitzak 64 lotura dauzka), kanpoko kable eta *transceiver* -en kostua eta energia-kontsumoa murriztuta. Horretaz gain, txasis bakoitza bi bideragailuz (B) osatuta dago: lotura kopuru aldakorra (*L*) daukatenak eta beste bideratzaile batzuen arteko konexioak egiteko erabiltzen direnak (*lotura* eta *ataka* terminoak berdin erabiliko ditugu hemendik aurrera). Irudian ikusi daiteke KE bakoitzak 32 lotura erabiltzen dituela bideratzaile bakoitzera konektatzeko. Bideragailuak FPGAak erabilia eginda daudenez, *L*-ren balioa aldatu daiteke bestelako ezaugarriak dauzkaten sareak sortzeko. Azken mailak armairua du izena (AR), eta 16 txasisez osatuta dago. Aipatu behar da kopuru hori alda daitekeela irudian erakusten den bezala, baina planifikatuta dagoen prototipoak 16 txasis erabiliko ditu. Sistema osoak gehienez 256 armairu eduki ditzake.

Txasisen arteko interkonexioak egiteko bi aukera daude: lehenengoan, bideratzaileen arteko konexio zuzenak erabiltzen dira, eta armairuen arteko konexioak onartzen dira. Bigarrenean, aldiz, bideratzaile osagarriak (komertzialak edo neurrira egindakoak) jartzen dira txasisen artean, zeharkako topologia deritzona sortzeko. Esan beharra dago hirugarren aukera bat posible dela, armairu berean konexio zuzenak eta armairu desberdinen artean zeharkako konexioak erabilia lortzen dena. Lan honetan, eredu errazte aldera, ingurunea mugatu da topologia zuzenak baino ez sortzeko, baina posible da ingurunea aldatzea edozein topologia sortzeko.

3. INTERKONEXIO-SAREEN EZAUGARRIEN OPTIMIZAZIOA

Atal honetan interkonexio-sareetarako desiragarriak diren ezaugarriak definituko ditugu. Horiek erabiliz, eredu bat eraikiko da, sare-topologiak sortzen lagunduko duena. Hurrengo zerrendan deskribatzen dira zeintzuk diren gako-ezaugarri horiek:

- **Hutsegite-tolerantzia.** Arazoen aurrean sistemak funtzionatzen jarraitzeko daukan gaitasuna da. Ezaugarri hori hobetzeko sare-topologiaren konektagarritasuna optimizatu behar da.
- **Sarearen errendimendua.** ExaNeSt-en azken helburua barne-komunikazioak erabiltzen dituzten aplikazio paralelo erraldoiak egikaritzea da. Hori dela eta, errendimendu handiko interkonexio-sarea izatea derrigorrezkoa da. Helburu hori lortzeko hainbat ezaugarri

beharrezkoak dira: diametro txikia, banda-zabalera altua, eta iturrien eta helburuen arteko bide-aniztasun handia, hain zuzen ere.

- **Hedapen- eta mantentze-kostua.** Sarearen kostua ere ezaugarri ezinbestekoa da, eta ahal bezain baxu mantentzea garrantzi handikoa da. Hori lortzeko, txasisak konektatzeko erabiltzen diren kableen eta *transceiver* -en kopurua gutxitu behar da. Horretaz gain, elementuen kopurua gutxitzen bada, energia-kontsumoa eta ustiapen-kostua ere murriztu egingo dira.

Ezaugarri horiek lortzeko diseinatzen ari garen sarearen parametroak puntuan jarri behar ditugu. Baina, ezaugarri guztiak batera optimizatzeko neurri bakarra ez dagoenez, jarraian ezaugarri guztiak hartzen dituen gutxieneko neurri-multzo bat azalduko da. Horrekin batera, hurrengo definizioetan optimizazio-prozesuan erabiliko diren helburu-funtzioak aurkezten dira (s -k soluzio bat adierazten du, eta G -k soluzio hori erabilita sortutako grafoa).

Erdibitze-zabalera (EZ) sare bat bi parte desberdinetan zatitzeko moztu behar den lotura kopurua da. Neurri honek oso ondo islatzen ditu bai sistemaren errendimendua bai sistemaren hutsegite-tolerantzia. Horregatik, lehenengo helburu-funtzioak sarea adierazten duen grafoaren erdibitzea maximizatuko du. Baina grafo baten erdibitzea kalkulatzeko NP-oso motako problema bat da eta, horregatik, Kernighan/Lin algoritmo heuristikoa erabiltzen da [10] hurbilketa bat kalkulatzeko.

$$f_1(s) : EZ(G) \tag{1}$$

Bide-aniztasunak nodo-pairen arteko bide disjuntuen kopurua adierazten du, errendimenduaren eta hutsegite-tolerantziaren adierazle ona izanda. Errendimenduari dagokionez, bide asko izateak sareko lehia gutxitzen du, eta, hori dela eta, banda-zabalera hobeto baliatzen da. Era berean, hutsegite-tolerantziari dagokionez, zenbat eta bide gehiago, orduan eta hutsegite gehiago behar dira sarea deskonektatzeko.

Neurri hori kalkulatzeko, grafoaren batez besteko bide disjuntuen kopurua (BBD) neurtzen da. s soluzioa emanda, $BBD(s, i, h)$ (Bide Disjuntu Laburrenak) funtzioa definitzen da, iturri eta helburu baten arteko bide disjuntu laburren multzoa itzultzen duena. Funtzio hori erabiliz BBD kalkulatu da.

$$f_2(s) : 2 \times \frac{\sum_{i=1}^{|G|} \sum_{j=i+1}^{|G|} |BBD(a_i, a_j)|}{(|G|+1) \times |G|} \tag{2}$$

non $|G|$ bideratzaileen kopurua den, $a_i, a_j \in G$ eta f_2 -k topologiaren bideratzaile-pare guztien arteko batezbesteko bide disjuntuen kopurua adierazten duen.

Konektatutako ataken kopurua, txasisak interkonektatzeko erabiltzen dena. Neurri honek kostuarekin eta energia-kontsumoarekin erlazio lineala dauka, eta, horregatik, ezin-bestekoa da neurri hori gutxitzea. Hala ere, murriztapen horrek ondorio kaltegarriak dauzka aurretik aipatutako neurrietarako.

Neurri hori soluzio baten lotura aktiboen (LA) kopurua neurtuz kalkulatzen da, hau da, txasisak konektatzeko erabiltzen diren bideratzaileen lotura kopurua. s soluzioa emanda, lotura aktiboen multzoa honela definitzen da: $LA^s = \{\exists i \in \{1, \dots, 2 \times R\} \text{ eta } s(i) > 0\}$ eta multzo hori erabiliz hirugarren funtzioa honela definitzen da:

$$f_3(s) : |LA^s| \tag{3}$$

Aipatutako hiru neurriak sarearen ezaugarri praktikoen adierazle esanguratsuak dira. Hori dela eta, lan honetan, sare-topologien optimizazioa gauzatuko da gatazkan dauden hiru neurriak erabilita.

4. HELBURU-ANIZTASUN OPTIMIZAZIO ALGORITMOAK

Atal honetan erabili ditugun algoritmoak azalduko dira helburu-aniztasuneko problema gisa formulatuta. Mota horretako algoritmo multzotik bi aukeratu ditugu: NSGA-II, eta SMS-EMOA hain zuzen ere.

Optimizazioaren pauso bakoitzean (belaunaldi deritzona), algoritmo bakoitzak gizabanakoen multzo bat (populazioa) gordetzen du. Multzo horretako elementuak problema baterako hautagai diren soluzioak dira. Soluzio baten kalitatea aurreko atalean azalduko funtzioak erabiliz neurtzen da. Eskuarki, pauso bakoitzean gizabanako onenak aukeratzen dira irizpide baten arabera, eta, horiek erabiliz, populazio berri bat (ondorengo) eraikiko da. Horietaz gain, populazio berriari gizabanako gehiago gehitu behar zaizkio, eta horiek momentuko populazioari gurutzatze-eragileak eta mutazio-eragileak aplikatuz sortzen dira. Pauso horiekk gelditzeko irizpidea bete arte errepikatzen dira.

Algoritmo horiek soluzio multzo bat sortzen dute helburu guztiak aldi berean optimizatuz. Helburu horiek lehian egon behar dute, hau da, helburu bat hobetzen denean besteren bat okerragotuko da. Horri gainartzaile kontzeptua deitzen zaio. Soluzio-multzoari Pareto-multzoa deitzen zaio eta zenbakizko balioak dituen multzoari Pareto-fronte [7]. Ondoren, formalki helburu-aniztasuneko optimizazio-problema bat definitzen da murrizketa batzuen menpe:

$$\min\{f_1(x), \dots, f_{N_{Hel}}(x)\} \tag{4}$$

$$\begin{cases} g_j(x) = 0 & j = 1, \dots, M, \\ h_j(x) \leq 0 & j = 1, \dots, K \end{cases} \tag{5}$$

non f_i i -garren helburu-funtzioa den, x soluzio bat den, N_{hel} helburuen kopurua den, $M + K$ murrizketen kopurua den eta g_j and h_j problemaren murrizketak diren. Jarraian, aukeratutako algoritmoak azalduko ditugu, eta baita haien arteko desberdintasuna ere.

Non-dominated Sorting Genetic Algorithm II: NSGA-II [7] algoritmoaren helburua populazio baten eta Pareto-fronte baten arteko ahokadura hobetzea da. Populazio hori azpi-populazioaren hierarkia batean ordenatzen da Paretoaren menpekotasuna erabiliz. Azpimultzo bakoitzako kideen arteko antzekotasuna neurtzen da Pareto-frontean, eta ondoriozko multzoak eta neurriak erabiltzen dira menpean ez dauden soluzioez osatutako multzo bat sortzeko.

S Metric Selection EMOA: SMS-EMOA algoritmoak hautagai diren soluzio onenak aukeratzeko *hypervolume* adierazlea (S neurria) [3] erabiliz. Neurri honek Pareto-menpekotasuna betetzen du; hau da, neurri honen balio altuena daukan multzoak bere mendera dauzka bestelako multzoak. Algoritmo honen bidez sortzen den populazioa oso banatuta dago, eta, horregatik, Pareto-frontearen leku interesgarrietan jartzen da arreta.

5. OPTIMIZAZIO PROBLEMAREN DEFINIZIOA

Lan honen helburua da optimizazioa erabiliz, bideratzaileen arteko loturen esleipenak sortzea. Hori lortzeko, optimizazio-prozesua hasten da ausazko gizabanakoak (lehenengo populazioa) sortzen, hiru helburu optimizatzeko erabiliko direnak: erdibitze-zabalera eta bide-aniztasuna maximizatzeko, eta, horiekin batera, loturen kopurua minimizatzeko. Atal honetan definituko da nola adierazten diren problema horren soluzioak eta algoritmo ebolutiboek nola erabiltzen dituzten gurutzatze-eragilea eta mutazio-eragilea.

5.1. Problemaren definizioa eta gizabanakoen adierazpidea

Topologia bat definitzeko hiru parametro erabiliko dira: $|L|$, lotura kopurua; $N \leq 16$, armairu bakoitzean dagoen txasis kopurua, eta $M \leq 256$, armairu kopurua. Bideratzaileen arteko loturen esleipen bat sortzeko loturen multzoan (L) bijekzio bat (φ) aurkitu behar dugu, lotura bakoitzari ($l \in L$) beste batekin ($l' \in L$) esleitzen diona.

$$\begin{aligned} \varphi : L &\in L \\ l &\mapsto \varphi(l) = l' \end{aligned}$$

Soluzio baten adierazpidea bi helburu kontuan hartuta sortzen da; alde batetik, definizioa errazteko, eta, bestetik, sistemaren erregularutasuna bermatzeko. Gure ereduak txasisen arteko distantzia erlatiboa (nodo

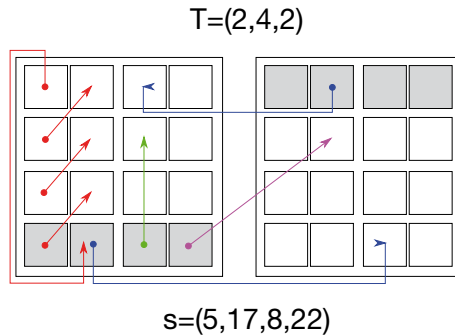
identifikatzailea erabiliz) definitzen du. Horrela, bi nodoren arteko distantzia txasis baten nodo kopurua baino txikiagoa izanez gero *barneko konexioa* izango da, eta txasis bereko nodoen arteko konexioak moduleragiketa erabiliz adieraziko dira. Bestela, *kanpo-konexioak* izango dira. Eredu horrek txasis baten loturak baino ez ditu adierazten, eta sistemaren gainerakoa simetria erabiliz eraikiko da. Adierazpide berria honelakoa da: $s = (s(1), \dots, s(2 \times R)) = (l'_1, l'_2, \dots, l'_2|L|)$ non

- $s(i) \in [(|L| \times 2) - i + 1, (N \times |L| \times 2) - i]$ armairu bereko lotura denean (*barneko konexioa*) edo
- $s(i) \in [(N \times |L| \times 2) - i + 1, (M \times N \times |L| \times 2) - i]$ armairu ezberdinen arteko lotura denean (*kanpo-konexioak*).

2. irudian, s gizabanakoa erabilita sortzen diren bideratzaileen arteko loturak erakusten dira, topologiaren parametroak $T = (|L| = 2, N = 4, M = 2)$ izanda. Barneko konexioak gezi gorriak eta berdeak erabiliz marraztu dira 5 eta $8 \leq (2 \times |L| \times N = 16)$ distantziara hurrenez hurren). Bestela, kanpo-konexioak dira 17 eta $22 > (2 \times |L| \times N = 16)$ distantziara daudelako hurrenez hurren, eta gezi urdinak eta arrosak erabiliz marraztu dira. 0 distantziara dauden konexioak erabili gabekotzat hartzen dira.

5.2. Helburu-funtzioak eta eragile espezifikokoak

Azpiatal honetan optimizatuko diren hiru helburu-funtzioak aurkezten dira, 3. atalean azaldu direnak. *EZ* eta *BBD* funtzioak emanda, (grafo baten erdibitzea eta bide disjuntuen kopurua itzultzen dutenak, hurrenez hurren), optimizazio-prozesuak maximizatuko ditu lehenengo eta bigarren helburu-funtzioak: $maxf_1 : EZ(G)$ eta $maxf_2 : BBD(G)$. Bestalde, hirugarren helburu-funtzioa, lotura aktiboen kopurua neurtzen duena, minimizatu egin behar da: $minf_3 : |LAs|$.



2. irudia. s gizabanakoa erabilita sortzen diren bideratzaileen arteko loturak.

Lehenago azaldu denez, helburu-aniztasuneko optimizazioan helburuek lehian egon behar dute eta horrela gertatzen da erabilitakoekin. Adibidez, banda-zabalera optimizatzeak gutxitu ditzake bide disjuntuen kopurua eta erabilitako lotura kopurua.

Erabili ditugun eboluzio-algoritmoek eragile hauek baliatzen dituzte:

- Gurutzatze-eragileak bi hautagai berri sortzeko (haurrak) bi gizabana (gurasoak) nahasten ditu. Prozesu hori gauzatzeko gurasoen konexio batzuk trukutzen dira.
- Mutazio-eragileak P_{mut} probabilitatearekin indibiduoak aldatzen ditu, edo konexio berri bat sartuz (P_{ber} probabilitateduna) edo jadanik dagoen konexio bat kenduz ($1 - P_{ber}$). Bi konexio mota desberdin daude: kanpo-konexioak (P_{kan} probabilitateduna) edo barne-konexioak ($1 - P_{kan}$ probabilitateduna).
- Gurutzatze- edo mutazio-eragilea erabilia indibiduo berri bat sortu ondoren, sortutako sare-topologia guztiz konektatuta dagoela ziurtatu behar da.

Erabilitako algoritmoek soluzio-bilduma bat sortzen dute, erdibitzearen, bide-aniztasunaren eta ataka-kopuruaren arteko batezbesteko desberdinak dituen. Gainera, sortutako soluzioak balioztatzeke analisi enpiriko bat egiten da guk garatutako INSEE simulatzailea erabiliz.

6. LAN ESPERIMENTALA

Atal honek egin den lan esperimental aurkezten du, eta bi galdero erantzuten saiatzen da: (1) zein optimizazio-algoritmok sortzen ditu zenbaki-emaizta onenak? eta (2) zenbat errendimendu ematen dute sortutako sare-topologiek?

1. taula. Optimizazio-algoritmoen parametroen konfigurazioa.

Parametroak	Balioa	Deskribapena
K_{pop}	100	Indibiduo kopurua belaunaldiko.
K_{bel}	100	Belaunaldi-kopurua.
P_{gur}	0,8	Gurutzatze-eragilearen probabilitatea.
P_{mut}	0,8	Mutazio-eragilearen probabilitatea.
P_{berria}	0,5	Mutazio-motaren probabilitatea.
P_{karp}	0,25, 0,5, 0,75	Kanpoko loturaren probabilitatea.

6.1. Optimizazio-algoritmoen ebaluazioa

Lehenik, begiratu diezaiegun helburu-funtzioen balioei. Hori egiteko, 10 esperimentu egin ditugu, zeinek 10 Pareto-multzo sortu dituzten. Emaitza horiek erakusteko, ataken kopuru berdina daukaten soluzioak batzen ditugu.

Egindako esperimentuetan erabili ditugu lau armairu, eta bakoitza 16 txasisez eta, gehienez, 16 irteera daukaten 16 bideratzailez osatuta dago: $-T = (8, 16, 4)$. Sortutako topologiengatik ezaugarriak aztertzeke, barne-konexioen eta kanpo-konexioen arteko proportzioa aldatzen denean, P_{kanp} hiru balio erabiltzen ditugu: 0,25, 0,50 eta 0,75. 1. taulan adierazi ditugu erabilitako parametroak. Kontuan izan behar da lan honen helburua ez dela parametro onenak bilatzea, eta, horregatik, parametro berdinak erabili ditugu algoritmo guztietan.

Sortutako topologiengatik zenbakizko emaitzen bidez, 2. taulan agertzen direnak, problema honetarako optimizazio-algoritmoen errendimendua konpara dezakegu. Gehienetan, NSGA-II algoritmoak f_1 eta f_2 baliorik altuenak lortzen ditu edozein f_3 baliotarako. Horretaz gain, p_{kanp} probabilitateak ia ez dauka eraginik algoritmoen portaeran; hau da, NSGA-II algoritmoak emaitza onenak lortzen ditu edozein kasutan.

Taula horren azken bi lerroetan Pareto-multzoa erabiliz sortutako topologiengatik batez besteko diametroa eta distantzia adierazi ditugu. Neurri horiek topologiengatik kalitatearen adierazle gisa erabil daitezke, nahiz eta optimizazio prozesuan parte ez hartu. SMS-EMOA algoritmoak lortzen ditu batez besteko distantzia eta diametro baxuenak kasu guztietan. Dena den, propietate horien balioak kontuan izanda, algoritmo bien portaera antzekoa dela ikus daiteke. Hala eta guztiz ere, balio horiek topologiengatik adierazle teorikoak baino ez direla nabarmendu nahi dugu, eta, horregatik, aplikazioetan izango duten egiazko eragina hurrengo atalean aztertuko da.

6.2. Topologiengatik ebaluazioa

Aurreko atalean sortutako soluzioak ebaluatuko ditugu INSEE simulatzailea [12] erabiliz, ingterkonexio-sareak aztertzeke askotan erabili baita. INSEEk bideratzaileen eredu oso zehatza dauka (phit mailan), eta bai trafikoko sintetikozko patroiak bai egiazko aplikazioetatik ateratako lan-kargak erabil ditzake. Horretaz gain, hainbat mekanismo dauzka elkar blokeatzea saihesteko topologia eta bideraketa askotarako [14]. Topologia horien errendimendua neurtzeko trafikoko uniforme eta *bisect* izeneko trafikoa erabili ditugu. Lehenengoak datu-paketeen helburuak ausaz sortzen ditu trafikokoaren banaketa uniforme sortzeko, eta, era berean, sarearen karga orekatzeko gaitasuna ulertzeko erabiltzen da. Bigarrenetan, aldiz, KEen erdiak

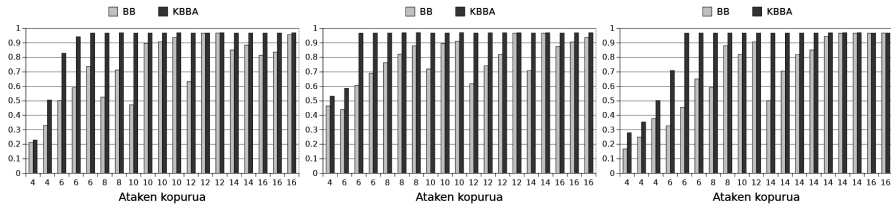
beste erdiekin komunikatzen dira eta erdibitze-zabalera neurtzeko erabiltzen da. Sistema osoak 128 KE dauzka, hau da, 4 armairu, 16 bideratzaile armairuko eta 2 KE bideratzaileko.

2. taula. NSGA-II eta SMS-EMOA algoritmoak erabilia lortutako f_1 eta f_2 funtzioen batez besteko balioak. Emaitzak f_3 erabiliz elkartu dira (ataken kopurua). Azken bi leurotan Pareto-multzo bakoitzaren batez besteko diametroa eta distantzia adierazi dira.

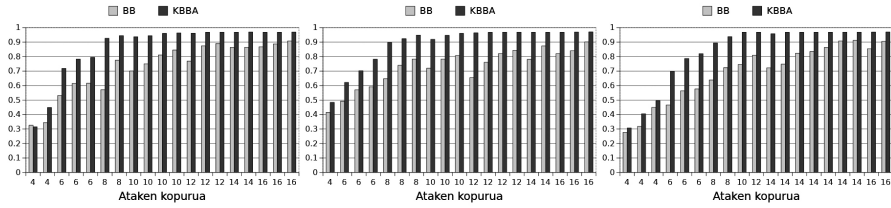
	$P_{kamp} = 0, 25$				$P_{kamp} = 0, 50$				$P_{kamp} = 0, 75$			
	NSGA-II		SMS-EMOA		NSGA-II		SMS-EMOA		NSGA-II		SMS-EMOA	
f_3	μf_1	μf_2	μf_1	μf_2	μf_1	μf_2	μf_1	μf_2	μf_1	μf_2	μf_1	μf_2
4	36,89	13,46	39,00	7,98	39,00	8,59	38,55	6,97	40,42	10,99	38,89	5,08
6	79,67	8,74	80,67	5,86	74,76	22,30	80,87	6,54	80,00	9,37	74,87	7,89
8	131,89	6,24	113,23	7,31	133,89	9,30	124,42	6,35	131,76	10,92	123,87	5,81
10	186,82	4,99	152,87	6,06	176,22	10,39	163,34	5,52	168,00	10,46	176,11	4,82
12	232,29	5,68	197,57	7,40	228,53	7,20	214,74	5,44	233,26	9,77	215,04	5,63
14	277,61	7,06	235,70	6,51	281,37	6,20	260,19	6,07	276,50	8,50	264,81	6,06
16	333,45	6,33	258,85	5,81	328,84	6,99	297,23	6,35	336,00	6,35	309,07	5,97
Diam.	5,89		5,33		6,70		5,98		6,49		5,59	
Dist.	3,46		3,20		3,85		3,53		3,78		3,32	

Errendimendu-konparazio bat egiteko auzazko topologia bat erabili da, *Jellyfish* hain zuzen ere, errendimendu handikoa eta sortu berria dena. Topologia honetan, sistema osoak 128 KE dauzka, hau da, 128 bideratzaile eta KE bat bideratzaileko. Bide-aniztasunaren optimizazioak errendimenduan daukan eragina ulertzeko, bi bideraketa-mota erabiltzen ditugu. Alde batetik, bi nodoren arteko bide bakarra (BB), laburrena, erabiltzen duen bideraketa ebaluatuko da. Bestetik, kostu bereko bide-aniztasuneko (KBBA) [1] bideraketa erabiltzen da, bi nodoren arteko bide laburrenak diren guztiak erabiltzen dituen.

Lehenengo esperimendu-multzoan (ikusi 3. irudia), $P_{kamp} = 0,25, 0,50$ eta $0,75$ probabilitateekin sortutako Pareto-multzoen errendimendua neurtuko da. Irudiaren goialdean trafiko uniformearen erabiliz sortutako topologiaren errendimendua irudikatu da. Espero genuen bezala, BB bideraketa erabiliz gero, ataken kopuruak eragin handia dauka errendimenduan, baina kontuan izan behar da kasu horietan energia-kontsumoa askoz baxuagoa izango dela. Hala ere, KBBA bideraketa erabiltzen bada, ikus daiteke errendimendua gehitzen dela bide gehiago erabiltzen baititu komunikazioak egiteko. Era berean, ikus dezakegu $P_{kamp} = 0,50$ probabilitatearekin sortutako topologiak, oro har, errendimendu haundiagoa lortzen dutela BB zein KBBA erabiliz.



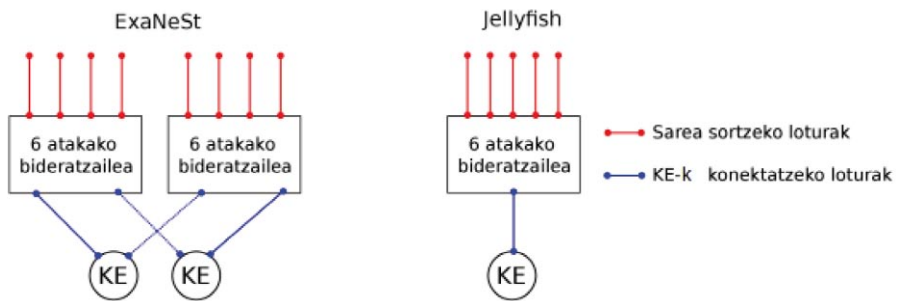
(a) 0,25 eta trafiko uniforme. (b) 0,50 eta trafiko uniforme. (c) 0,75 eta trafiko uniforme.



(d) 0,25 eta *bisect* trafikoa. (e) 0,50 eta *bisect* trafikoa. (f) 0,75 eta *bisect* trafikoa.

3. irudia. Hiru Pareto-multzoen soluzioak erabiliz sortutako sare-topologiaren errendimendua, BB eta KBBA bideraketak erabiltzen dituztenak.

Orain *bisect* trafikoa erabiliz lortutako emaitzak neurtuko ditugu (ikus 3d, 3e eta 3f. irudiak). Trafiko hori erabiliz atera ditzakegun ondorioak antzekoak dira, eta berriro ikus daitezke KBBAk errendimenduan daukan eragina oso handia dela. Horretaz gain, emaitzek argi eta garbi erakusten dute ataken kopurua gehituz gero topologiaren erdibitze-zabalera handituko dela.



4. irudia. Konputazio-elementuak bideratzaileetara konektatzeko era. Irudian, 6 atakako bideratzaileak erabiltzen dituen konfigurazioa ikus daiteke.

Bigarren esperimentu-multzoan, sortutako topologiak konparatu ditugu errendimendu handiko sare batekin. Horretarako, zazpi konfigurazioa

zio erabili ditugu. Adibidez, 1. konfigurazioa sortzeko (ikus 4. irudia), lau ataka erabiltzen dituen soluzio bat ateratu da $p_{kamp} = 0,25$ probabilitatearekin sortutako pareto-multzo batetik, eta, era berean, bi soluzio gehiago $p_{kamp} = 0,50$ eta $p_{kamp} = 0,75$ probabilitateekin sortutako pareto-multzoetatik. Konfigurazio horiek sei atakako bideratzaileak erabiltzen dituzte, lau sare sortzeko eta beste biak bi KE konektatzeko. Era berean, *Jellyfish* sareak sei atakako bideratzaileak erabiltzen ditu, baina, kasu horretan, bost ataka erabiliz sortzen da sarea, eta KE bat baino ez dago konektatuta bideratzaile bakoitzera. Beste konfigurazio batzuk ere era berean sortu ziren 8, 10, 12, 14, 16 eta 18 atakako bideratzaileak erabiliz. Emaitzak 5. irudian irudikatu ditugu.

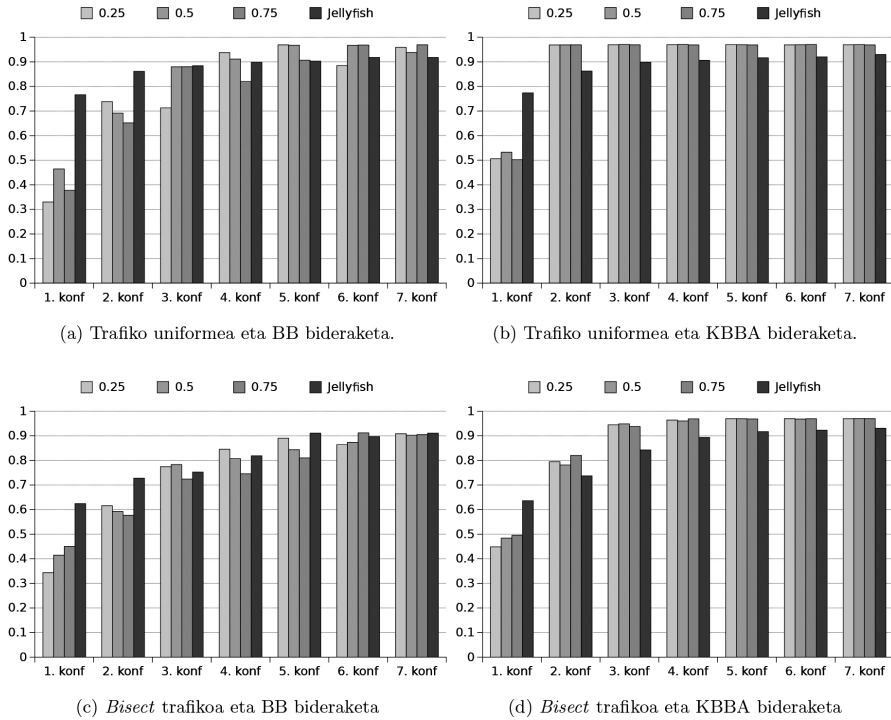
Trafiko uniforme eta BB bideraketa erabiliz lortutako errendimendua aztertzen hasiko gara (ikus 5a. irudia). Argi eta garbi ikus daitezke lehenengo hiru konfigurazioekin *Jellyfish*-ek lortzen duela errendimendu altuena. Hala ere, beste konfigurazioak erabiltzen direnean, optimizazio bidez sortutako topologiek errendimendu altuagoa lortzen dute. Bigarren emaitza-multzoari erreparatu badiogu, KBBA bideraketa erabiltzen duena, ikus dezakegu gure inguruneak sortutako topologiak hobereak direla, lehenengoan izan ezik. Portaera horren arrazoia da bideraketa horrek topologiaren bide-aniztasuna aprobetxatzen duela.

Bukatzeko, *bisec* trafikoa erabiliz sortutako emaitzak aztertuko ditugu, 5c. eta 5d. irudietan daudenak. Kasu horretan, *Jellyfish* ek lortzen du errendimendu altuena BB bideraketa erabiltzen denean. Hala ere, KBBA bideraketa erabiltzen denean, berriro egoera guztiz aldatzen da, eta, lehenengo konfigurazioan izan ezik, inguruneak sortutako topologiek errendimendu altuagoa lortzen dute.

7. ONDORIAK ETA ETORKIZUNERAKO LANA

Lan honetan eskala handiko super-konputagailuetarako sare-topologiak automatikoki sortzeko optimizazio-ingurune bat aurkeztu da. Ingurune hori ExaNeSt proiektuan erabiliko da eraikitzen ari garen sistemaren sarea garatzeko. Ingurune horrek eskaintzen duen malgutasuna oso baliagarria da gure proiekturako, ezaugarri bereziak optimizatzea eta tamaina desberdinetako bideratzaileak erabiltzea ahalbidetzen baitigu.

Gure inguruneak helburu-aniztasuneko optimizazio algoritmo ezagunetarikoa bi erabiltzen ditu: NSGA-II eta SMS-EMOA, hain zuzen ere. Gainera, bideratzaileen arteko konexioak adierazteko era berri bat sortu da, bilaketa-espazioa murrizten laguntzen duena eta sare-topologia erregularrak sortzen dituena. Horretaz gain, gurutzatze- eta mutazio-eragile espezifikoak garatu ditugu problema honetarako.



5. irudia. Inguruneak sortutako sare-topologiaren eta *Jellyfish* sare-topologiaren arteko errendimendu konparazioa.

Emaitzek, sistema txikiak erabilia, argi eta garbi adierazten dute NSGA-II-k SMS-EMOAK baino soluzio hobek sortzen dituela. Gainera, simulazioen bidez egindako esperimenduek adierazten dute, oro har, inguruneak sortutako sare-topologiak *Jellyfish* topologiak baino errendimendu altuagoa lortzen dutela, batez ere KBBA bideraketa erabiltzen denean.

Lan honetan aurkeztu den ingurunea sare-topologiaren sortzaile bat garatzeko lehenengo urratsa da. Gure asmoa da, alde batetik, optimizazio-algoritmoen kopurua gehitzea (esaterako, berriki garatutako NSGA-III [16]), eta, bestetik, ereduaren kopurua gehitzea bestelako sare-topologiak sortzeko (adibidez, sare irregularrak eta hibridoak). Amaitzeko, aipatu nahi dugu sistema askoz handiagoak erabilia lortutako emaitzak aurkeztu nahi ditugula.

8. ESKER ONAK

Ikerketa hau finantzatua izan da Europar Batasuneko ExaNeSt (No 671553) proiektuaren eta Erresuma Batuko (EPSRC) »Interconnection Networks: Practice unites with Theory (INPUT)»(EP/K01568 0/1 eta EP/K015699/1) proiektuaren bitartez.

9. ERREFERENTZIAK

- [1] Analysis of an Equal-Cost Multi-Path Algorithm. RFC 2992, Nov. 2000.
- [2] ARM. <https://www.arm.com>.
- [3] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective Selection based on Dominated Hypervolume. *European Journal of Operational Research*, 181(3):1653-1669, 2007.
- [4] D. Chen, N. A. Eisley, P. Heidelberger, R. M. Senger, Y. Sugawara, S. Kumar, V. Sala-pura, D. L. Satterfield, B. Steinmacher-Burow, and J. J. Parker. The IBM Blue Gene/Q Interconnection Network and Message Unit. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1-10, New York, NY, USA, 2011. ACM.
- [5] C. Concatto, J. A. Pascual, J. Navaridas, J. Lant, A. Attwood, M. Lujan, and J. Goodacre. A CAM-free Exascalable HPC Router. In *31st International Conference on Architecture of Computing Systems*, ARCS'18. ACM, 2018.
- [6] Cray Inc. http://www.cray.com/assets/pdf/products/xe/idc_948.pdf.
- [7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182-197, 2002.
- [8] M. Katevenis et al. The ExaNeSt Project: Interconnects, Storage, and Packaging for Exascale Systems. In *2016 Euromicro Conference on Digital System Design (DSD)*, 00:60-67, 2016.
- [9] R. Ammendola et al. The next generation of exascale-class systems: The exanest project. In *2017 Euromicro Conference on Digital System Design (DSD)*, pages 510-515, 2017.
- [10] B. W. Kernighan and S. Lin. An Efficient Heuristic Procedure for Partitioning Graphs. *The Bell System Technical Journal*, 49(2):291-307, Feb 1970.
- [11] J. Kim, W. J. Dally, S. Scott, and D. Abts. Technology-Driven, Highly-Scalable Dragonfly Topology. In *Proceedings of the 35th International Symposium on Computer Architecture*, pages 77-88, Washington, 2008.
- [12] J. Navaridas, J. Miguel-Alonso, J. A. Pascual, and F. J. Ridruejo. Simulating and evaluating interconnection networks with insee. *Simulation Modelling Practice and Theory*, 19(1):494-515, 2011.
- [13] ORNL. <https://www.olcf.ornl.gov/summit/>.
- [14] J. A. Pascual and J. Navaridas. High-Performance, Low-Complexity Deadlock Avoidance for Arbitrary Topologies/Routings. In *ACM International Conference on Supercomputing, ICS '18. ACM/SIGARCH*, 2018.

- [15] A. Singla, C.-Y. Hong, L. Popa, and P. B. Godfrey. Jellyfish: Networking Data Centers Randomly. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, NSDI'12, pages 17-17, Berkeley, CA, USA, 2012.
- [16] Y. Yuan, H. Xu, and B. Wang. An Improved NSGA-III Procedure for Evolutionary Many-objective Optimization. In *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*, GECCO'14, pages 661-668, New York, NY, USA, 2014. ACM.