REF2021

Institution: University of Westminster	
Unit of Assessment: 11 Computer Science and Informatics	
Title of case study: Component-Based Message Passing in Java for Distributed Computing	
Period when the underpinning research was undertaken: 2000 - 2020	
Details of staff conducting the underpinning research from the submitting unit:	

Name(s): Vladimir Getov	**Role(s) (e.g. job title):** Professor of Distributed and High-Performance Computing	**Period(s) employed by submitting HEI:** Since 22 Feb 1995

Period when the claimed impact occurred: 1 August 2013 – 31 December 2020	
Is this case study continued from a case study submitted in 2014? ~~Y~~/N	

1. Summary of the impact (indicative maximum 100 words)

Vladimir Getov's pioneering research on message passing for Java (MPJ) provided a crucially important framework and programming environment for parallel and distributed computing with Java. This invention resulted in an industry standard specification and a novel MPJ-based hierarchical development methodology for a new generation of large-scale grid and cloud-based distributed systems. These achievements led to:

- *Impact on Professional Practice*: Getov's MPJ specification provided the foundations for the Java binding in Open MPI (message passing interface) – the most popular worldwide open-source library, which provides easy access to this binding for application programmers.
- *Economic Impacts*: IBM's biometric identification system, built in collaboration with Getov and using the MPJ framework, has had significant economic impact for IBM via the quicker return of investment achieved by its higher productivity and shorter development cycle. The adoption of Getov's MPJ development methodology by ActiveEon has also seen the company demonstrate significant revenue growth (18M USD across the REF period).
- *Social Impacts*: IBM Watson's development was based on Getov's component composition message passing approach and has demonstrable benefits in the areas of social services, environmental protection, and public safety. ActiveEon's use of Getov's MPJ development methodology in their ProActive workflow and scheduling framework has had significant benefits in biomedical distributed applications.

2. Underpinning research (indicative maximum 500 words)

Released in 1995, the Java programming language was rapidly adopted by industry and end users because of its portability and internet programming support. However, Java did not have the symmetric message passing capability, widely recognised as vitally important for parallel and distributed memory computing. By contrast, efficient message passing support had already been captured in the MPI standard for other programming languages such as C, C++, and Fortran.

Professor Vladimir Getov identified this weakness and conducted pioneering research and development on MPJ which has been a main area of focus for the Distributed and Intelligent Systems Research Group (DIS RG) at the University of Westminster since the late 90s. This research team comprises: Professor V. Getov, Dr A. Bolotov, Dr S. Isaiadis, Dr S. Minchev, Dr T. Weigold, Dr A. Basukoski, Dr J. Thiyagalingam, and Dr A. Basso. The early success of DIS RG attracted high international interest and the creation of the International MPJ Working Group with participation from the UK, USA, Europe, and Japan.

Chaired by Vladimir Getov, the MPJ Working Group developed a methodology for building mixed-language MPJ applications which evolved from three approaches: (a) wrapping of existing MPI libraries via hand-written software; (b) automatic code generation of the wrapper interface by a novel tool-generator; and (c) development from scratch of the MPI libraries in Java. Getov has participated in the development of all three approaches by implementing the MPJ specification and ensuring full compatibility with the already existing and very successful MPI standard. The **MPJ results successfully resembled MPI, providing symmetric message passing for distributed computing with Java** leading to its adoption by the professional community [1].

This working group was then invited to join the Global Grid Forum (an international professional organisation focusing on grid and cloud computing) where Getov furthered this research by expanding the application of MPJ into modern large-scale distributed systems such as grids and

clouds, both of which allow for the analysis of huge data sets. Tackling the scalability and productivity challenges, Getov and colleagues formally specified building blocks called "components" and **introduced component-based approaches that enabled MPJ to provide the interconnection mechanisms in complex grid and cloud systems** [2].

Continuing work in this area, the next main research challenge was to replace the Common Component Architecture (CCA) model. Available in early 2000, the CCA provided only limited support for a single coupling of components in a 2-dimensional space. Getov's innovation was to develop a novel abstract approach for hierarchical (multiple) composition of components in a 3-dimensional space. Further, Getov invented a novel MPJ-based hierarchical components composition development methodology for a new generation of large-scale grid and cloud-based distributed systems. This provided the theoretical background for a recursive and efficient component-based development methodology. The combination of these research contributions to the European CoreGRID project significantly advanced this new field through Getov's **development of the hierarchical Grid Component Model (GCM) specification for various distributed computing systems** [3]. Together with his DIS RG team, Getov followed up the GCM specification with proof-of-concept experiments in a development environment that used the hierarchical components MPJ methodology and provided confidence about the higher efficiency of this novel approach [4].

Building on this work, the DIS RG team, led by Getov, was a main partner in the European GridComp project. Working in close collaboration with other partners including INRIA Sophia Antipolis (France), IBM-Research (Switzerland), Atos Origin (Spain), University of Pisa (Italy), and Tsinghua University (China), the GridComp project designed and built a fully functional platform incorporating the ICBE (Integrated Component-Based Environment) prototype for hierarchical components composition [5]. A major contribution of GridComp was **the design and implementation of a component based MPJ framework for rapid development and deployment of efficient grid and cloud computing applications**. This work led to **four new international standards**, approved by ETSI (European Telecommunications Standards Institute): "GCM Interoperability Deployment", Aug 2008; "GCM Interoperability Application Description", Aug 2008; "GCM Fractal ADL", Mar 2009; "GCM Management API (Java, C, WSDL)", Mar 2010.

Further results on smart cloud architectures followed the expanded ICBE methodology, which included **three approaches for developing GCM applications** [6]. Getov contributed significantly to each: The first, a wrapper approach for legacy codes reuse, supports both hand-written and automatically generated code. The second approach componentises existing applications via appropriate modifications. The third approach is component-based development from scratch. This work confirmed **the significantly higher productivity of the component-based MPJ development methodology**.

---

3. References to the research (indicative maximum of six references)

1. B. Carpenter, V. Getov, G. Judd, A. Skjellum, G. Fox. 2000. MPJ: MPI-like Message Passing for Java, *Concurrency: Practice and Experience*, vol. 12 (11), pp. 1019-1038. DOI: 10.1002/1096-9128(200009)12:11<1019::AID-CPE518>3.0.CO;2-G
2. V. Getov, G. von Laszewski, M. Philippsen, I. Foster. 2001. Multi-Paradigm Communications in Java for Grid Computing, *Communications of the ACM*, vol. 44(10), pp. 118-125. DOI: 10.1145/383845.383872
3. F. Baude, D. Caromel, C. Dalmasso, M. Danelutto, V. Getov, L. Henrio, C. Pérez. 2009. GCM: A Grid Extension to Fractal for Autonomous Distributed Components, *Annals of Telecommunication*s, vol. 64(1-2), pp. 5-24. DOI: 10.1007/s12243-008-0068-8
4. A. Basukoski, V. Getov, J. Thiyagalingam, S. Isaiadis. 2008. Component-Based Development Environment for Grid Systems: Design and Implementation, In: *Making Grids Work*, Springer, pp. 119-128. DOI: 10.1007/978-0-387-78448-9_9
5. T. Weigold, M. Aldinucci, M. Danelutto, V. Getov. 2012. Process-Driven Biometric Identification by means of Autonomic Grid Components, *Int. J. of Autonomous and Adaptive Communications Systems*, vol. 5(3), pp. 274-291. DOI: 10.1504/IJAACS.2012.047659
6. V. Getov. 2011. Component-Oriented Approaches for Software Development in the Extreme-Scale Computing Era, In: *High Performance Computing: From Grids and Clouds to Exascale*, IOS Press, pp. 141-156. DOI: 10.3233/978-1-60750-803-8-141

Funding in GBP (selected):
• CoreGrid: Software Infrastructures and Applications for Large-scale Distributed, Grid and Peer-to-Peer Technologies, 224,000 (Total consortium grant 7.4 M) • Research Fellowship, European Commission, 14,100. • Visiting Lab Fellow, Pacific Northwest National Laboratory, 37,200. • GridComp: Grid Programming with COMPonents, 320,000 (Total consortium grant 2.7 M) • Smart Cloud Infrastructures – IBM Faculty Award, 10,000. • ComplexHPC: High Performance Computing in Complex Environments, 16,710 (Total consortium grant 470 K)

4. Details of the impact (indicative maximum 750 words)

Impact 1: Professional Practice **- Open Source and Standardisation**
Since August 2014, the MPJ specification has been included in the core distribution of the widely used Open MPI (Open-Source Message Passing Interface) software environment. Developed and maintained by a consortium of research, academic, and industry partners, Open MPI is a suite of open-source software libraries implementing the MPI standard for High Performance Computing (HPC) [a-i]. The Java binding developed in Open MPI is based on the ground-breaking results of the International MPJ Working Group chaired by Vladimir Getov (output [1]) and incorporate the four ETSI standards that resulted from the GridComp project (outputs [4], [5] and [6]). The Open MPI team have highlighted that these **'standardized classes' of Java are of benefit to an open-source library** as they 'provide a platform-independent way to access host-specific features such as threads, graphics, file management, and networking' [a-ii, p.4 & 2]. In other words, these standardised classes guarantee full compatibility and portability across a wide variety of platforms and applications.

In recent years, Open MPI has become the most popular MPI library, as is evidenced by the MPI International Survey, which received over 800 user responses from 42 different countries between Feb & July 2019 [a-iii, p.1]. Of the MPI users identified, 85.8% (718 respondents) use Open MPI [a-iv]. As such, **through its inclusion in this open-source library, MPJ has been reaching a wide range of users**, 80% of which belong to non-profit organisations, including government institutes, as well as software and hardware vendors, with the use of Java being especially popular in the UK and Russia [a-i, p.13 & 16].

Regarding the significance of this wide range of users, the Open MPI team explain that MPJ was added as it is one of the three key parallel computing technologies upon which HPC depends and has several benefits. For instance, it 'provides efficient built-in support for threads' which increases the software development productivity. In addition, 'some numerical libraries are based on multithreading' and MPI itself 'can benefit from Java because its widespread use makes it likely to find new uses beyond traditional HPC applications' [a-ii, p.1-2]. As such **the primary impact on professional practice relates to providing easy access to Java binding for programmers** who can exploit these benefits (output [2]). Secondary impacts arise through specific examples of such usage of MPJ via Open MPI. For instance, the PPF library for parallel particle filtering applications allows application programmers to write quickly and easily shared- and distributed-memory PPF codes in Java [a-v], producing the benefit of '**reducing the cost of traditional particle filters** by approximating the likelihood with a mixture of uniform distributions over pre-defined cells or bins.'

Impact 2. Economic Impact
An example of the significant economic impact on users of Prof Getov's innovative component based MPJ development methodology is its contribution to 'the long-term research and development collaboration for IBM's Cloud Computing agenda and strategy', through knowledge exchange activities at IBM Research Centers in Zurich, Almaden, Watson, and Dublin, as well as direct collaborative research [b]. Dr Peter Buhler, Head of Computer Science at the Zurich Laboratory states that this has resulted in their 'using and developing further the adopted component-based methodology and development platform which has [in turn] influenced the professional practice in building complex applications for grid and cloud systems' at IBM [b].

For instance, Buhler [b] confirms that **IBM Watson** was 'designed by reusing the principles developed and standardized by the European Telecommunications Standards Institute as part of Vladimir Getov's work' (output [5]). This supercomputer is a unique artificial intelligence system

capable of answering questions posed in natural language and its economic significance to IBM is indicated by its global market share of 16.1% placing it amongst the three dominant systems in the Machine Learning category [d-i]. IBM Watson has been used by nearly 2,800 companies as of July 2020 due to its usefulness across a variety of application domains. Statistics by revenue show that 39% of Watson users are large companies (>$1000M), 37% are small (<$50M), and 16% are medium-sized, and that they cut across different sectors of industry – the largest being Computer Software (22%), Hospital & Health Care (14%), Higher Education (8%), and Information Technology and Services (7%) [d-i].

Buhler also cites **'the Biometric Identification System application** developed in collaboration with Vladimir's DIS RG team and delivered by IBM Research – Zurich Laboratory' (Fig. 1), as having a 'significant economic impact' through its producing 'quicker return on investment […] achieved because of the much higher productivity and shorter development cycle provided by the invented component-based methodology and development process' [b]. This biometric identification system contributes to the Global Technology Services segment of IBM's business, which in 2019 comprised the largest share of IBM's total revenue at nearly 36% ($27.4 billion) [c-ii]. The use of Getov's component based MPJ framework provided 'pre-built template components' from which a system can be built that 'guarantee[d] real-time biometric identification functionality over a very large, constantly growing database of enrolled identities (fingerprints)' [b]. Furthermore, Getov's ICBE (outputs [4], [5] and [6]) 'has been instrumental in continuously monitoring and maintaining the system' [b], which has been a significant aspect of IBM's commercial offering to corporations and governments [c-i].

Fig. 1: Security Biometric Identification System [c-iii]
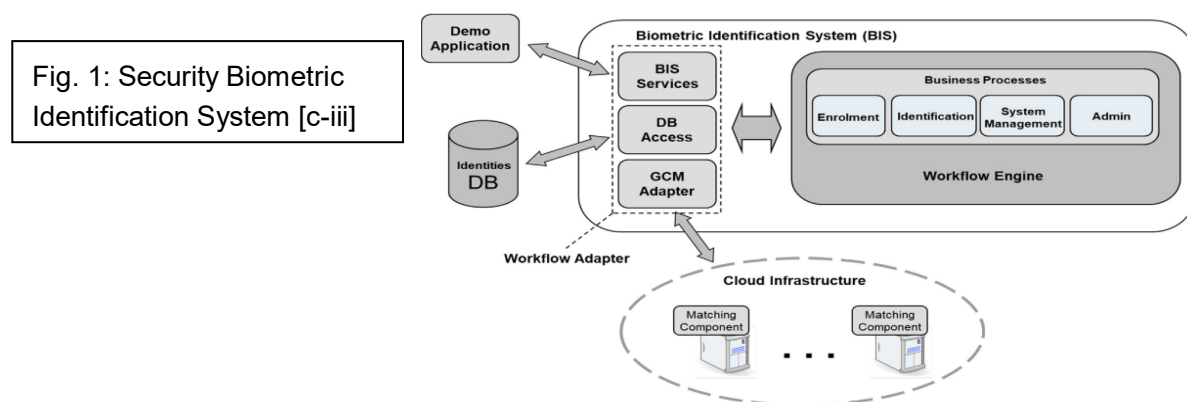
Another company that has economically benefitted from Getov's innovative component-based MPJ methodology, as well as the underpinning ETSI standards, is ActiveEon – a France-based software company that provides innovative open-source solutions for IT automation, acceleration and scalability, big data, distributed computing and application orchestration. During the GridCOMP project (06/2006 – 02/2009), Getov and colleagues had initially demonstrated the usefulness of the component based MPJ development methodology to industry by using it to wrap and Grid-enable aerodynamic wing modelling software at ActiveEon, and to prove the integration of data staging for the input and output files into this sweeping / optimisation process for any given configuration.

As Prof Denis Caromel, CEO of ActiveEon, states, since then the company 'has been directly exploiting the results of the project, including the GCM and MPJ framework (output [3]), particularly in its *ProActive workflows and scheduling open-source* middleware which has been used for delivering solutions for a range of commercial customers' [e-i]. Caromel further specifies the economic impact of this adoption of Getov's innovations: 'over the years since August 2013, **ActiveEon has achieved revenue growth of 18M USD starting with less than 1M USD in 2013 and reaching revenue of 19M USD in 2020**. Our number of employees has been increasing significantly over the same period from 30 employees at the end of 2013 to 106 employees in 2020. Those successful results would not have been possible without Professor Getov's direct contribution to the component based MPJ development methodology and ETSI standards adopted and used by ActiveEon throughout the 13 years of its existence. His work has reduced substantially the return of investment cycle of complex distributed applications' [e-i].

Impact 3. Social Impact

Beyond the significant economic impact Vladimir Getov has made with his contributions, Buhler explains that '[t]he technical advancements he [Getov] has invented and co-invented are the basis for many Web-scale applications we have at our fingertips today', creating a 'resulting impact on the society' [b]. IBM specify three key social impacts of their cognitive technology in lengthy case studies featuring real-world usage of the IBM Watson supercomputer that would not exist without Getov's research [d-ii to d-iv]. The areas of interventions listed below provide typical examples:

• **Transforming social services** by providing 'easier-to-access, more personalized services [that] can help individuals at risk better manage their own well-being, getting the right support when they need it' [d-ii]. A specific use case is 'Aspiranet, which currently serves 22,000 youth and families across California, [and] is using natural language inquiry of unstructured data to help youth transition from foster home care to living on their own'; their CEO confirms 'cognitive technology helps free up caseworkers' time, enabling them to focus on what matters most, human connection' [d-ii]. This is a significant impact given that in the 'United States, social worker turnover is as high as 90 percent per year, with heavy workloads, including paperwork, as a major contributing factor' [d-ii].

• **Providing environmental protections:** IBM gives the example of their work with the Beijing Environmental Protection Bureau to reduce air pollution in China (the cause of more than one million premature deaths per year) via the 2014 Green Horizons initiative, which 'addresses these challenges by using advanced machine learning to identify smaller sections of the city that are at risk. Along with trade-off analyses, this enables more targeted mitigation actions, such as shutting down targeted industries, while minimizing socioeconomic disruptions' [d-iii].

• **Enhancing public safety**: IBM gives the example of how they used Watson to anticipate and respond to the Oct 2015 Hurricane Patricia (one of the strongest storms ever recorded): 'weather prediction models based on artificial intelligence (AI) provided advanced warning to an IBM production center in Guadalajara. The system had analyzed huge volumes of disparate information, including weather data, social feeds and news reports to get a comprehensive view of the storm's trajectory. […] Early warning gave site officials the crucial time they needed to act' and they 'opted to evacuate the site as a precautionary measure' [d-iv].

Another example of social impacts that have been created through the adoption of Getov's innovations is found in the work of ActiveEon. As mentioned earlier, ActiveEon has used Getov's component based MPJ development methodology in their *ProActive Workflows & Scheduling* software which incorporates dynamic and autonomous workload allocation to an elastic pool of resources using multi-language software components within the HPC cloud-based workflows [e-i] [e-ii, p.30]. Among a range of case studies demonstrating usage of this software across sectors including mining, satellite technology, visa processing, and solvency solutions [e-iii], is its use for microbiome analytics orchestration [e-iv, p.18-20]. The enabling of microbiome analytics orchestration at this scale is of **significant benefit to the biomedical field as microbiome innovation encompasses pharmaceutical applications** such as personalised medicine and nutrition, identification of the impact of drugs on the gut microbiome, and identification of bacterial biomarkers associated with dysbiosis (microbial imbalance) [e-iv, p.8].

5. Sources to corroborate the impact (indicative maximum of 10 references)

[a] (i) Open MPI: Open Source High Performance Computing [site] (ii) O. Vega-Gisbert et al. 2016. Design and implementation of Java bindings in Open MPI. *Parallel Computing 59*: 1–20 [link] (iii) A Report of MPI International Survey, *EuroMPI/USA*, 2020 (iv) MPI International Survey, March 2019 [link] (v) PPF Library [link]

[b] Dr Peter Buhler, Distinguished research staff member and Head of the Computer Science Department at the IBM Research Lab – Zurich; testimonial letter [PDF]

[c] (i) IBM Identity and Access Management [site] (ii) IBM revenue sources [link] (iii) GridCOMP, 'Advanced Grid/Cloud Programming with Components: A Biometric Identification Case Study'

[d] (i) Companies using IBM Watson, *enlyft* data analysis [link]; Social Impacts of IBM Watson: (ii) Social services [Link], (iii) Environment [link], (iv) Public safety [link]

[e] (i) Prof Denis Caromel, founder and CEO of ActiveEon – testimonial letter [PDF] (ii) ActiveEon's ProActive Workflows: from HPC to Data Analytics to Machine Learning [link] (iii) ActiveEon use cases [link] (iv) Microbiome Analytics: Machine Learning ActiveEon Orchestration On-Prem and On-Clouds [link]